Models of functional neuroimaging data

Stephan, K E ; Mattout, J ; David, O ; Friston, K J

Abstract: Inferences about brain function, using functional neuroimaging data, require models of how the data were caused. A variety of models are used in practice that range from conceptual models of functional anatomy to nonlinear mathematical models of hemodynamic responses (e.g. as measured by functional magnetic resonance imaging, fMRI) and neuronal responses. In this review, we discuss the most important models used to analyse functional imaging data and demonstrate how they are interrelated. Initially, we briefly review the anatomical foundations of current theories of brain function on which all mathematical models rest. We then introduce some basic statistical models (e.g. the general linear model) used for making classical (i.e. frequentist) and Bayesian inferences about where neuronal responses are expressed. The more challenging question, how these responses are caused, is addressed by models that incorporate biophysical constraints (e.g. forward models from the neural to the hemodynamic level) and/or consider causal interactions between several regions, i.e. models of effective connectivity. Some of the most refined models to date are neuronal mass models of electroencephalographic (EEG) responses. These models enable mechanistic inferences about how evoked responses are caused, at the level of neuronal subpopulations and the coupling among them.

DOI: https://doi.org/10.2174/157340506775541659

Published in final edited form as:
Curr Med Imaging Rev. 2006 February ; 2(1): 15–34.

Models of functional neuroimaging data

Klaas Enno Stephan[*], **Jeremie Mattout**, **Olivier David**, and **Karl J. Friston**
The Wellcome Dept. of Cognitive Neurology, University College London Queen Square, London, UK WC1N 3BG

Abstract

Inferences about brain function, using functional neuroimaging data, require models of how the data were caused. A variety of models are used in practice that range from conceptual models of functional anatomy to nonlinear mathematical models of hemodynamic responses (*e.g.* as measured by functional magnetic resonance imaging, fMRI) and neuronal responses. In this review, we discuss the most important models used to analyse functional imaging data and demonstrate how they are interrelated. Initially, we briefly review the anatomical foundations of current theories of brain function on which all mathematical models rest. We then introduce some basic statistical models (e.g. the general linear model) used for making classical (i.e. frequentist) and Bayesian inferences about *where* neuronal responses are expressed. The more challenging question, *how* these responses are caused, is addressed by models that incorporate biophysical constraints (e.g. forward models from the neural to the hemodynamic level) and/or consider causal interactions between several regions, i.e. models of effective connectivity. Some of the most refined models to date are neuronal mass models of electroencephalographic (EEG) responses. These models enable mechanistic inferences about how evoked responses are caused, at the level of neuronal subpopulations and the coupling among them.

Keywords

fMRI; EEG; MEG; modelling; statistical inference; dynamic systems; effective connectivity

I Introduction

Imaging neuroscience depends on conceptual, anatomical, statistical and causal (i.e. neurobiologically and biophysically motivated) models that link ideas about how the brain works to observed neuronal or hemodynamic responses. The aim of this review is to demonstrate the relationships among the different models that are used in modern neuroimaging. We will show how simple statistical models, used to identify *where* evoked brain responses are expressed can be elaborated to provide models of *how* neuronal responses are caused (e.g. models of effective connectivity). These successive elaborations rely, increasingly, on biological mechanisms. We will review a series of models that cover conceptual models, motivating experimental design, to detailed biophysical models of coupled neuronal ensembles that enable the researcher to address rather complex questions about neurophysiological and computational processes. Note that we will not discuss any physical or methodological foundations of the various imaging modalities that this review refers to. Readers who are unfamiliar with the principles of functional magnetic resonance imaging (fMRI), positron emission tomography (PET), magnetoencephalography (MEG) or

[*] Corresponding author: Tel (44) 020 7833 7485 Fax (44) 020 7813 1445 E-mail: k.stephan@fil.ion.ucl.ac.uk.

electroencephalography (EEG) are referred to standard textbooks of imaging neuroscience, e.g. [1].

The structure of this paper and the conceptual relations between the various models discussed are summarized in Figure 1. Anatomically motivated theories of functional brain architectures represent the fundaments of neuroimaging. In Section II we start by reviewing the distinction between *functional specialisation* and *functional integration* and how these principles serve as the basis for most models of neuroimaging data. In section III, we turn to simple statistical models (e.g. the general linear model) used for making classical and Bayesian inferences about functional specialisation, in terms of *where* neuronal responses are expressed. Characterising a region-specific effect rests on estimation (of models parameters) and inference (about the magnitude of these parameters). Inferences in neuroimaging may concern differences between groups of subjects or changes within subjects over a sequence of observations. They may pertain to structural differences (e.g. in voxel-based morphometry [2]) or to neurophysiological indices of brain functions (e.g. fMRI or EEG). This paper is only concerned with inferences about functional phenomena. We will initially focus on the analysis of fMRI time-series, because the relevant models cover most of the issues encountered in other modalities. By incorporating biological constraints, simple observation models can be made more realistic and, in a dynamic framework, rendered causal. This section concludes by considering some of the recent advances in biophysical modelling of hemodynamic responses. All the models considered in this section pertain to regional responses. In section IV, we focus on models of distributed responses, where the interactions among cortical areas or neuronal subpopulations are modelled explicitly. This section covers the distinction between *functional connectivity* and *effective connectivity* and focuses on one of the most recent approaches, Dynamic Causal Modelling (DCM), based on fMRI and EEG data. We conclude with an example from ERP (event-related potential) research and show how the P300 can be explained by changes in coupling among neuronal sources that may underlie perceptual learning.

II Anatomical models

1. Functional specialisation and functional integration

The functional organisation of the brain seems to obey two main principles, *functional specialisation* and *functional integration,* where the integration within and among specialised areas is mediated by effective connectivity. Functional localisation implies that a function can be localised in a cortical area, whereas specialisation suggests that a cortical area is specialised for some aspects of cognitive processing, and that this specialisation is anatomically *segregated* within the cortex. The cortical infrastructure supporting a single function may then involve many specialised areas whose union is mediated by the functional integration among them. In this view functional specialisation is only meaningful in the context of functional integration and *vice versa* [3].

From a historical perspective, the distinction between functional specialisation and functional integration relates to that between *localisationism* and *[dis]connectionism* that dominated thinking about brain function in the nineteenth century. Since the formulation of phrenology by Gall who postulated fixed one-to-one relations between particular parts of the brain and specific mental attributes, the identification of a particular brain region with a specific function has become a central theme in neuroscience. Somewhat ironically, the notion that distinct brain functions could, at least to some degree, be localised in the brain, was strengthened by early scientific attempts to refute the phrenologists' claims. In 1808, a scientific committee of the Athénée at Paris, chaired by Cuvier, declared that phrenology was an unscientific and invalid theory [4]. This conclusion, which was not based on experimental results, may have been enforced by Napoleon Bonaparte (who, allegedly, was not amused after Gall's phrenological examination of his own skull did not give the flattering results he expected). During the

following decades, however, this strong verdict triggered the development of lesion and electrical stimulation procedures for the experimental investigation of animal brains to test whether cognitive functions could indeed be localised. Initial lesion experiments by Flourens on pigeons gave results incompatible with phrenologist predictions, but later experiments, including stimulation experiments in dogs and monkeys by Fritsch, Hitzig and Ferrier, supported the idea that there was a relation between distinct brain regions and certain cognitive or motor functions. Additionally, clinicians like Broca or Wernicke showed that patients with focal brain lesions in particular locations showed very specific cognitive impairments. However, it was realised early on that, in spite of these experimental findings, it was generally difficult to attribute a specific function to a cortical area, given the dependence of cerebral activity on the anatomical connections between distant brain regions [5]. For example, although accepting the results of electrical stimulation in dog and monkey cortex, Goltz considered that the excitation method was inconclusive, in that movements elicited might have originated in related pathways, or current could have spread to distant centres [6]. Some years later, observations on patients with brain lesions that affected fibre tracts led to the concept of *disconnection syndromes* and the refutation of localisationism as a complete or sufficient explanation of cortical organisation [7].

2. Functional specialisation and functional segregation

On the basis of theoretical considerations and analyses of connectivity patterns, it has been argued that the functional role of any brain unit (e.g. cortical area, subarea or neuronal population) is defined largely by its connections [8,9]. Moreover, certain patterns of cortical projections are so common that they could amount to rules of cortical connectivity. “These rules revolve around one, apparently, overriding strategy that the cerebral cortex uses - that of functional segregation” [10]. Functional segregation demands that cells with common functional properties be grouped together. This architectural constraint necessitates both convergence and divergence of cortical connections. Extrinsic connections among cortical regions are not continuous but occur in patches or clusters. This patchiness has, in some instances, a clear relationship to functional segregation. For example, area V2 has a distinctive cytochrome oxidase architecture, consisting of thick stripes, thin stripes and inter-stripes. When recordings are made in V2, directionally selective (but not wavelength or colour selective) cells are found exclusively in the thick stripes. Labelling of V2 cells after injections of retrograde tracer in area V5 is limited to these thick stripes [11]. All the available physiological evidence suggests that V5 is a functionally homogeneous area that is specialised for visual motion. Evidence of this nature supports the notion that patchy connectivity is the anatomical infrastructure that mediates functional segregation and specialisation. If it is the case that neurons in a given cortical area share a common responsiveness, by virtue of their extrinsic connectivity, to some sensorimotor or cognitive attribute, then this functional segregation is also an anatomical one.

In summary, the concept of functional specialisation suggests that challenging a subject with the appropriate sensorimotor attribute or cognitive process should lead to activity changes in, and only in, the specialised areas. This is the anatomical and physiological model upon which the search for regionally specific effects, pursued by means of neuroimaging techniques, is based. We will deal first with models of regionally specific responses and return to models of functional integration later.

III Statistical models of regional responses

1. Statistical parametric mapping

Functional mapping studies are usually analysed with some form of statistical parametric mapping. Statistical parametric mapping entails the construction of spatially extended

statistical processes to test hypotheses about regionally specific effects [12]. Statistical parametric maps (SPMs) are image processes with values for each volume element (voxel) that are, under the null hypothesis, distributed according to a known probability density function, usually the Student's T or F distributions. These are known colloquially as T- or F-maps and often referred to as SPM{T} and SPM{F}, respectively. The success of statistical parametric mapping is due largely to the simplicity of the idea. Namely, one analyses each and every voxel using a standard (univariate) statistical test. These usually test for activation, or regression on some explanatory variable. The resulting statistical parameters are assembled into an image - the SPM. SPMs are interpreted as statistical processes that are continuous in space (or sometimes time) by referring to the probabilistic behaviour of random fields [12-14]. Random fields model both the univariate probabilistic characteristics of a SPM and any non-stationary spatial covariance structure under the null hypothesis. 'Unlikely' excursions of the SPM are interpreted as regionally specific effects, attributable to the sensorimotor or cognitive process that has been manipulated experimentally.

Over the years statistical parametric mapping [15] has come to refer to the conjoint use of the *general linear model* (GLM) and *random field theory* (RFT) to analyse and make classical inferences about spatially extended data through statistical parametric maps. The GLM is used to *estimate* some parameters that could explain the spatially continuous data in exactly the same way as in conventional analysis of discrete data. RFT is used to resolve the multiple-comparisons problem that ensues when making *inferences* over a multitude of voxels contained by the brain volume analysed (the "search volume"). RFT provides a method for adjusting p-values for the search volume of an SPM to control false positive rates. It plays the same role for statistical tests on spatially continuous data as the Bonferroni correction for a family of discrete statistical tests.

Later we will consider the Bayesian alternative to classical (i.e. frequentist) inference with SPMs. This rests on conditional inferences about an effect, given the data, as opposed to classical inferences about the data, given the effect is zero. Bayesian inferences about effects that are continuous in space use Posterior Probability Maps (PPMs; [16]). Although less established than SPMs, PPMs are potentially very useful, not least because they do not have to contend with the multiple-comparisons problem induced by classical inference (see [17]). In contradistinction to SPM, this means that inferences about a given regional response do not depend on inferences about responses elsewhere. Before looking at the models underlying Bayesian inference we first consider estimation and classical inference in the context of the GLM.

2. The general linear model (GLM)

Statistical analysis of imaging data corresponds to (i) modelling the data to partition observed neurophysiological responses into components of interest, confounds and error and (ii) making inferences, about interesting effects, using the variances of the partitions. A brief review of the literature may give the misleading impression that there are numerous ways to analyse PET and fMRI time-series, with a diversity of statistical and conceptual approaches. With few exceptions, however, every analysis is a variant of the GLM. Different types of analyses that are all derived from the general linear model include: (i) simple T-tests on PET scans assigned to one condition or another, (ii) correlation coefficients between observed responses and boxcar stimulus functions in fMRI, (iii) analyses using multiple linear regression, (iv) analysis of (co) variance, (v) selective averaging to estimate event-related responses, and (vi) linear time-invariant convolution models of evoked responses. Mathematically, they are all identical and can be implemented with the same equations and algorithms. The only thing that distinguishes among them is the design matrix encoding the experimental design.

The general linear model is an equation

$$y=X\beta+\varepsilon \quad (1)$$

expressing the observed response y, at each voxel, in terms of a linear combination of explanatory variables in the matrix X plus a well-behaved error term ε. The matrix X that contains the explanatory variables (e.g. designed effects or confounds) is called the *design matrix*. Each column of the design matrix corresponds to some effect that has been experimentally manipulated or that may confound the results. These are referred to as explanatory variables, covariates or regressors. These explanatory variables encode experimental effects that are assumed to be expressed in a linear and instantaneous fashion in the data, without reference to any particular mechanism. The design matrix commonly contains indicator variables or parametric variables encoding the experimental manipulations. These are formally identical to classical ANOVA or multiple linear regression models, respectively [18].

Each column of the design matrix X has an associated but unknown parameter that has to be estimated. Some of these parameters, which are assembled in the vector β, will be of interest (e.g. the effect that a particular cognitive condition has on the magnitude of the neurophysiological response). The remaining parameters will be of no interest and pertain to nuisance or confounding effects (e.g. signal drifts over time or head movements). The statistical test is directed to interesting effects by specifying the null hypothesis with a *contrast*. A contrast is simply a linear mixture of parameter estimates. The T-statistic allows one to test the null hypothesis that some contrast (e.g. the difference between conditions) of the estimates is zero. The T-statistic obtains by dividing the contrast (specified by contrast weights) of the parameter estimates, by its standard error (Figure 2). Sometimes, several contrasts are tested jointly, for example, when using polynomial [19] or basis function [20] expansions of some experimental factor. In these instances, the F statistic is used. An F-contrast is specified with a matrix of contrast weights that can be thought of as a collection of T-contrasts that one wants to test en masse.

It is straightforward to apply the GLM to PET data because PET experiments are designed such that each scan can be assigned to a particular condition and long inter-scan periods make individual observations (i.e. scans) independent of each other. In contrast, as described in the following paragraphs, additional issues need to be considered in the case of fMRI and EEG because here the observations correspond to time series.

As for fMRI, this technique does not measure instantaneous neural responses but a relatively sluggish hemodynamic signal, the so-called “blood oxygen level dependent” (BOLD) response which is typically sampled at a rate between 0.25-1 Hz. This has two major consequences. First, while an experimentally controlled event evokes a transient neural response almost instantaneously, the associated BOLD response follows with a few seconds delay and is dispersed in time. This means that the BOLD responses to experimental events which are close in time overlap and superimpose in some fashion. A GLM can take these features of BOLD responses into account by means of a hemodynamic impulse response function (HRF). The HRF describes the characteristic hemodynamic response to a brief neural event and thus characterizes the input-output behavior of a given voxel. In the standard convolution model for fMRI, each voxel is treated as an independent linear time-invariant (LTI) system, and the explanatory variables (e.g. stimulus functions) are convolved with a canonical HRF to give predicted hemodynamic responses that enter the design matrix as regressors [21]. This convolution model will be discussed in more detail in section 5.1.

Second, unlike PET scans, successive fMRI scans are not independent. This induces temporal autocorrelation among the errors in Eq. 1. This is a problem when making inferences about the

model parameters because the classical T- and F-statistics assume the errors to be i.i.d. (independently and identically distributed) and thus the error covariance matrix to be a multiple of the identity matrix. Any departure from this assumption is referred to as "non-sphericity". Two methods exist to deal with non-sphericity due to temporal autocorrelations: "pre-colouring" and "pre-whitening". In both cases, Eq. 1 is multiplied with a filter matrix S to give

$$Sy=SX\beta+S\varepsilon \quad (2)$$

(see Figure 2). With pre-colouring, one tries to replace the unknown endogenous autocorrelation by imposing a known autocorrelation structure (i.e. using a pre-defined *S*). This known autocorrelation can then be taken into account when making inferences by using a generalized least squares scheme (see [22] for details). In contrast, pre-whitening tries to estimate *S* from the data such that the errors in Eq. 2 are uncorrelated (i.e. become white noise) and the error covariance matrix becomes proportional to the identity matrix [23,24]. Pre-whitening is, in principle, more efficient whereas pre-colouring is less prone to bias in estimating the standard error of the parameters [25].

In contrast to analyses of PET and fMRI data, the application of GLMs to EEG data is a relatively recent development. Kiebel & Friston [26,27] have suggested a general framework in which the mass-univariate approach of SPM developed initially for fMRI and PET can be applied to event-related potentials (ERP). This approach has proven useful at either the sensor (scalp) or reconstructed source (cortical) level. The particular type of questions asked of ERP data (e.g. inference about differential latencies among conditions), means that time needs to be considered as an experimental factor rather than simply as a fourth dimension of the data.

2.1 Classical Inference—Having estimated the parameters and computed the chosen statistic, RFT is used to assign adjusted *p*-values to topological features of the SPM, such as the height of peaks or the spatial extent of regions above a threshold. This *p*-value is a function of the search volume and smoothness of the residuals [14]. The intuition behind RFT is that it allows one to control the false positive rate of peaks or "blobs" (i.e. clusters of voxels above a certain threshold) corresponding to regional effects. A Bonferroni correction would control the false positive rate of voxels, but this is inexact and unnecessarily severe because it neglects the smoothness of the data, i.e. spatial dependencies that exist among voxels. If the RFT-adjusted *p*-value for a particular regional effect (and thus the probability of observing this effect by chance) is sufficiently small (usually less than 0.05), the regional effect can be declared significant.

The equations for the general linear model, non-sphericity correction and inferential statistics summarised in Figure 2 can be used to implement a vast range of analyses. The issue is therefore not so much the mathematics but the formulation of a design matrix *X* appropriate to the study design and inferences that are sought. Before considering general linear models as biophysical or causal models of brain responses we will focus on the design matrix as a device to specify experimental design, without reference to any particular mechanism how the data were caused.

3. Experimental design

This section considers the different sorts of designs employed in neuroimaging studies. Experimental designs can be classified as *single factor* or *multifactorial* designs, and within this classification the levels of each factor can be *categorical* or *parametric*.

3.1 Categorical designs, cognitive subtraction and conjunctions—The tenet of cognitive subtraction is that the difference between two tasks can be formulated as a separable cognitive or sensorimotor component. Regionally specific differences in the responses, evoked

by the two tasks, identify the corresponding functionally specialised area. Early applications of subtraction range from the functional anatomy of word processing [28] to functional specialisation in extrastriate cortex [29]. The latter studies involved presenting visual stimuli with and without some sensory attribute (e.g. colour, motion, etc.). The areas highlighted by subtraction were identified with homologous areas in monkeys that showed selective electrophysiological responses to equivalent visual stimuli.

Cognitive conjunctions [30] can be thought of as an extension of the subtraction technique, in the sense that they combine a series of subtractions. In subtraction ones tests a *single* hypothesis pertaining to the activation in one task relative to another. In conjunction analyses *several* contrasts are tested, asking whether all the activations, in a series of task pairs, are expressed conjointly. Consider the problem of identifying regionally specific activations due to a particular cognitive component (e.g. object recognition). If one can identify a series of task pairs whose differences have only that component in common, then the region which activates, in all the corresponding subtractions, can be associated with the common component. In other words, conjunction analyses allow one to disclose context-invariant regional responses.

3.2 Parametric designs—The premise behind parametric designs is that regional physiology will vary systematically with the degree of cognitive or sensorimotor processing, or deficits thereof. Examples of this approach include early PET experiments that demonstrated significant correlations between hemodynamic responses and the performance of a visually guided motor tracking task [31] or showed a clear linear relationship between perfusion in peri-auditory regions and frequency of aural word presentation [32]. This correlation was not observed in Wernicke's area, where perfusion appeared to correlate, not with the discriminative attributes of the stimulus, but with the presence or absence of semantic content. These relationships or *neurometric functions* may be linear or nonlinear. Using polynomial regression, in the context of the GLM, one can identify nonlinear relationships between stimulus parameters (e.g. stimulus duration or presentation rate) and evoked responses. To do this one usually uses a SPM{F} (see [19]).

The example provided in Figure 3 illustrates both categorical and parametric aspects of design and analysis. These data were obtained from an fMRI study of visual motion processing using radially moving dots [33]. Isoluminant and isochromatic stimuli, respectively, were presented over a range of speeds. To identify areas involved in visual motion a stationary dots condition was subtracted from conditions with moving dots (see the contrast weights on the upper right). To ensure significant motion-sensitive responses, under different levels of both colour and luminance, a conjunction of the equivalent subtractions was assessed under both viewing contexts. The resulting SPM{T} shows areas V5 and V3a. The T-values in this SPM are simply the minimum of the T-values for each subtraction. Thresholding this SPM ensures that all voxels survive a threshold in each subtraction separately. This *conjunction* SPM has an equivalent interpretation; it represents the intersection of the excursion sets, defined by the threshold of each *component* SPM. This intersection is the essence of a conjunction.

The responses in left V5 are shown in the lower panel of Figure 3 and speak clearly to an inverted ‘U’ relationship between speed and evoked response that peaks at around six degrees per second. It is this sort of relationship that parametric designs try to characterise. Interestingly, the form of these speed-dependent responses was similar using both stimulus types, although luminance cues are seen to elicit a greater response. From the point of view of a factorial design there is a *main effect* of cue (isoluminant vs. isochromatic), a [nonlinear] *main effect* of speed, but no speed by cue *interaction*.

3.3 Multifactorial designs—Factorial designs, where two or more factors are combined in the same experiment, are more prevalent than single factor designs because they enable

inferences about interactions. An interaction between two factors describes how the effects over the levels of one factor depends on the level of the other factor. Expressed simply, an interaction therefore represents a difference in a difference. Factorial designs have a wide range of applications. An early application, in neuroimaging, examined physiological adaptation and plasticity during motor performance, by assessing time by condition interactions [34]. Factorial designs have an important role in the context of cognitive subtraction and additive factor logic by virtue of being able to test for interactions, or context-sensitive activations (i.e. to demonstrate the fallacy of "pure insertion", [35]). Depending on the nature of the experimental factors, these interaction effects can sometimes be interpreted as (i) the integration of two (or more) cognitive processes or (ii) the modulation of one process by another.

To summarise this section on experimental design, the design matrix encodes the potential causes of observed data and, in particular, designed effects caused by changes in the level of various experimental factors. These factors can have categorical or parametric levels, and most experiments nowadays use multiple factors to test for both main effects and interactions. Before turning to mechanistically more informed formulations of the general linear model, we will consider briefly the two sorts of inferences that can be made about the parameter estimates.

4. Classical and Bayesian inference

To date, inference in neuroimaging has been restricted largely to classical (i.e. frequentist) inference based upon statistical parametric maps, using T or F statistics as described above. For each voxel, these SPMs can be used to compute the probability of the data under the null hypothesis that a particular effect or activation is absent. As the magnitude of an effect or activation is represented by some parameter β of the model[1], this probability of obtaining the data, given that the null hypothesis is true, can be written as $p(y|\beta=0)$. If this probability is sufficiently small (e.g. less than 0.05), the null hypothesis can be rejected and an inference is made that the activation is present.

The alternative approach is to use Bayesian or conditional inference based upon the posterior distribution of the parameters [16,24]. Given data y and parameters β, Bayes theorem states that the posterior distribution of the parameters $p(\beta|y)$ is proportional to the product of the likelihood $p(y|\beta)$ and the prior $p(\beta)$:

$$p(\beta|y) = \frac{p(y|\beta)\, p(\beta)}{p(y)} \tag{3}$$

A useful way to summarise this posterior density for a particular contrast of parameters $c^T\beta$ is to compute the probability that the contrast exceeds some threshold. This represents a Bayesian inference about the magnitude of the activation represented by the contrast, in relation to the specified threshold. By computing the posterior probability for each voxel we can construct posterior probability maps (PPMs) that are a useful complement to classical SPMs. The motivation for using Bayesian inference is that it has high face validity. This is because the inference is about an effect, or activation, being greater than some specified magnitude that has some meaning in relation to the underlying neurophysiology. This contrasts with classical inference, as described above, in which the inference is about the effect being significantly different from zero. The problem with this is that trivial departures from the null hypothesis can be declared significant, with sufficient data or sensitivity. From the perspective of neuroimaging, Bayesian inference is especially useful because it eschews the problem of multiple comparisons. In classical inference about neuroimaging data, which has the same

[1]More generally, effects of interest are represented by *contrasts* of parameters $c^T\beta$.

specificity for all voxels, one tries to ensure that the probability of rejecting the null hypothesis incorrectly is maintained at a small rate, despite making inferences over large volumes of the brain. This induces a multiple-comparisons problem that, for spatially continuous data, requires a correction to the *p*-value using RFT as described above, or alternatives like the False Discovery Rate (FDR; [36]). This correction means that classical inference becomes less sensitive or powerful with increasing search volumes. In contrast, Bayesian inference does not have to contend with the multiple-comparisons problem because the probability that an activation has occurred, given the data and a chosen threshold, at any particular voxel is the same, irrespective of whether one has analysed that voxel alone or the entire brain. This is achieved by estimating voxel-specific prior covariances for the parameters and thus adjusting the voxel-wise specificity of inference to ensure that it pertains to effects of the same size (see [16] and section 3 of [39] for details). This Bayesian perspective is similar to that of the frequentist who makes inferences on a per-comparison basis (see [17]). In conclusion, Bayesian inference using PPMs represents a relatively more powerful approach than classical inference in neuroimaging.

4.1 Hierarchical models and empirical Bayes—PPMs require the posterior distribution of the activation given the data, i.e. a contrast of conditional parameter estimates. This posterior density can be computed using Bayes rule. As shown by Eq. 3, Bayes rule requires the specification of a likelihood function and the prior density of the model's parameters. Under Gaussian assumptions, the models used to form PPMs and the likelihood functions are exactly the same as in classical SPM analyses, namely the GLM. The only extra bit of information that is required is the prior probability distribution of the parameters. Although it would be possible to specify this using independent data or some plausible physiological constraints, there is an alternative to this fully Bayesian approach. The alternative is *empirical Bayes* in which the prior distributions are estimated from the data. Empirical Bayes requires a hierarchical observation model where the parameters and hyperparameters[2] at any particular level can be treated as priors on the level below. For example, mixed-effects analyses of multi-subject studies are based on a two-level hierarchical model [37]. However, in neuroimaging there is a natural hierarchical observation model that is common to all brain mapping experiments. This is the hierarchy induced by looking for the same effects at every voxel within grey matter. The first level of the hierarchy corresponds to the experimental effects at any particular voxel and the second level comprises the effects over voxels. Put simply, the variation in a contrast, over voxels, can be used as the prior variance of that contrast at any particular voxel.

Linear hierarchical models have the following general form

$$\begin{aligned} y &= X^{(1)}\beta^{(1)} + \varepsilon^{(1)} \\ \beta^{(1)} &= X^{(2)}\beta^{(2)} + \varepsilon^{(2)} \\ \beta^{(2)} &= \dots \end{aligned} \tag{4}$$

The first line is exactly the same as Eq(1), but now the parameters of the first level are generated by a supraordinate linear model and so on to any hierarchical depth required. These hierarchical observation models are an important extension of the GLM. Parameters and hyperparameters in these hierarchical models are usually estimated using Expectation Maximisation (EM) [38,39]. In the context of PPMs (where one uses the variance of a contrast over voxels as the prior variance of that contrast at any particular voxel), y comprises the responses at all voxels, and $\beta^{(1)}$ are the experimental effects that we want to make an inference about. Because we

[2]Simply speaking, hyperparameters are parameters of the distribution of parameters.

have invoked a second level, the first-level parameters embody random effects and are generated by a second-level linear model:

$$\begin{aligned} y &= X^{(1)}\beta^{(1)} + \varepsilon^{(1)} \\ \beta^{(1)} &= 0 + \varepsilon^{(2)} \end{aligned} \quad (5)$$

At the second level of this model, $\beta^{(2)}$ is the average effect over voxels and $\varepsilon^{(2)}$ is its voxel-to-voxel variation. As the parameters of interest, $\beta^{(1)}$, reflect regionally specific effects, it is valid to assume that they sum to zero over all voxels. This corresponds to using a shrinkage prior (i.e. zero mean) at the second level; the variance of this prior is implicitly estimated by estimating the variance of $\varepsilon^{(2)}$. This empirical prior can then be used to estimate the posterior probability of $\beta^{(1)}$ being greater than some threshold at each voxel. An example of the ensuing PPM is provided in Figure 4 in comparison with the classical SPM.

Another application of hierarchical models, in an empirical Bayesian framework, is to solve the so-called EEG inverse problem to reconstruct the cortical current density which caused the scalp measures. Priors are indispensable for solving the inverse problem because the forward model is highly under-determined (the number of voxels or cortical sources is much larger than the number of measurement points). One approach uses a two-level model in the form of Eq. 5 where the data correspond to measured sensor time series from a time window of interest [40]. At the first level, the parameters $\beta^{(1)}$ represent the cortical source amplitudes, and the design matrix $X^{(1)}$ corresponds to the forward operator which defines the propagation of electric potentials through tissue to the sensors. At the second level, $X^{(2)} = 0$ so that the parameters of interest $\beta^{(1)}$ become a random variable with zero mean and a variance equal to the variance of $\varepsilon^{(2)}$. This variance is unknown but can be hyperparameterised as a linear combination of prior variance components, each of them corresponding to some prior information (constraint) on the underlying source distribution. Then, Restricted Maximum Likelihood (ReML) estimates of the hyperparameters associated with each prior variance component can be obtained using EM. Since the variance of the source parameters is estimated from the data, irrelevant or redundant priors can be included without loss of performance, provided that other informative priors are taken into account. This flexibility of empirical Bayes is of particular interest in multimodal integration, typically when using fMRI data to constrain EEG source reconstruction.

In this section we have demonstrated how the GLM can be used to test hypotheses about brain responses and how, in a hierarchical form, it enables empirical Bayes. In the next section we will deal with dynamic systems and how they can be formulated as GLMs. These dynamic models take us closer to how brain responses are actually caused by experimental manipulations and represent the next step toward causal models of brain responses.

5. Convolution models and temporal basis functions

Friston et al. estimated the form of the HRF using a least squares deconvolution [21]; the HRF can be used in linear-time invariant (LTI) models where neuronal responses, evoked by experimentally controlled stimulus functions, are convolved with an HRF to give a predicted hemodynamic response (see also [41]). This simple linear convolution model is the cornerstone for modelling activations in fMRI with the GLM. An impulse response function is the response to a single impulse, measured at a series of times after the input. It characterises the input-output behaviour of the system (i.e. voxel) and places important constraints on the sorts of inputs that will excite a response in that system.

Knowing the forms that the HRF can take is important for several reasons, not least because it allows for better statistical models of the data. The HRF may vary from voxel to voxel and

this has to be accommodated in the GLM. One option is to use temporal basis functions [20, 42,43]. The basic idea behind temporal basis functions is that the hemodynamic response, induced by any given trial type, can be expressed as the linear combination of several functions of peristimulus time. The standard convolution model for fMRI responses takes a stimulus function, encoding the neuronal responses, and convolves it with an HRF to give a regressor that enters the design matrix. When using basis functions, the stimulus function is convolved with all the basis functions to give a series of regressors (in Figure 2 we used four stimulus functions and two basis functions to give eight regressors). Mathematically we can express this model as

$$\begin{aligned} y(t) &= u(t) \otimes h(t) \\ h(t) &= \beta_1 T_1(t) + \beta_2 T_2(t) + \ldots \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} y(t) &= X\beta + \varepsilon \\ X_i &= T_i(t) \otimes u(t) \end{aligned} \tag{6}$$

where ⊗ means convolution. This equivalence illustrates how temporal basis functions allow one to take any convolution model (left) and convert it into a GLM (right). $u(t)$ is the experimental stimulus function for a particular trial type. The parameters β_i are the coefficients or weights that determine the mixture of basis functions of time $T_i(t)$ that best models $h(t)$, the HRF for the trial type and voxel in question. In the GLM, the convolution of the different basis functions T_i with the stimulus function defines the columns X_i of the design matrix X (see Eq. 6). We find the most useful basis set to be a canonical HRF and its derivatives with respect to the key parameters that determine its form (see below). Temporal basis functions are important because they enable a graceful transition between conventional multi-linear regression models with one stimulus function per condition and finite impulse response (FIR) models with a parameter for each time point following the onset of a condition or trial type. Figure 5 illustrates this graphically. In short, temporal basis functions offer useful constraints on the form of the estimated response that retain the flexibility of FIR models and the efficiency of single regressor models. In addition to temporal basis functions and FIR models, there are alternative approaches to model voxel-specific differences in the form of the HRF; see [44,45] for examples.

6. Biophysical models

6.1 Input-state-output systems—By adopting a convolution model for brain responses in fMRI we are implicitly positing some underlying dynamic system that converts neuronal responses into observed hemodynamic responses. Our understanding of the biophysical and physiological mechanisms that underpin the HRF has grown considerably in the past few years, and there now exist biophysically detailed and empirically validated models of the BOLD response (see [46,47] for recent reviews). The right panels in Figure 6 show results from simulations based on the Balloon model of BOLD responses [48] as extended by Friston et al. [49]. Here, neuronal activity, induced by an experimentally controlled stimulus function $u(t)$, triggers an auto-regulated vasodilatory signal (s) that causes transient increases in regional cerebral blood flow (f). This enhancement in blood flow dilates a venous balloon, increasing its volume (v) and diluting venous blood to decrease deoxyhemoglobin content (q). The BOLD signal is roughly proportional to the concentration of deoxyhemoglobin (q/v) and follows the change in flow with about one second delay (see Fig. 5). The model is framed in terms of differential equations, which are summarized in the left panel of Figure 6.

In this model we have introduced variables like blood flow and deoxyhemoglobin concentrations that are not actually observed directly. These are the *hidden states* in the *input-state-output model* described above. The general state and output equations of such a dynamical system are

$$\begin{aligned} \dot{x}(t) &= f(x,u,\theta) \\ y(t) &= g(x,u,\theta)+\varepsilon \end{aligned} \tag{7}$$

where $\dot{x}=\partial x/\partial t$. The first line is an ordinary differential equation and expresses the rate of change of the states as a function of the states $x(t)$, the inputs $u(t)$, and some time-invariant system parameters θ. Note that this model is deterministic, i.e. there is no process noise at the level of the hidden states. As in the examples above, the inputs $u(t)$ correspond to designed experimental effects (e.g. the stimulus function in fMRI). There is a fundamental and causal relationship [50] between the outputs and the history of the inputs in Eq. 7. This relationship conforms to a Volterra series, which expresses the output $y(t)$ as a generalised convolution of the input $u(t)$, critically without reference to the hidden states $x(t)$. This series is simply a functional Taylor expansion of the outputs with respect to the inputs [51]. The reason it is a *functional* expansion is that the inputs are a function of time:[3]

$$\begin{aligned} y(t) &= \sum_i \int_0^t \cdots \int_0^t \kappa_i(\sigma_1,\ldots,\sigma_i)\, u(t-\sigma_1),\ldots,u(t-\sigma_i)\, d\sigma_1,\ldots,d\sigma_i \\ \kappa_i(\sigma_1,\ldots,\sigma_i) &= \frac{\partial^i y(t)}{\partial u(t-\sigma_1),\ldots,\partial u(t-\sigma_i)} \end{aligned} \tag{8}$$

where $K_i(\sigma_1,\ldots\sigma_i)$ is the ith order kernel. In Eq. 8 the integrals are restricted to the past which renders Eq. 8 causal. The key thing here is that Eq. 8 is simply a convolution and can be expressed as a GLM in the same fashion in which we have described expansions by temporal basis functions as a GLM in Eq. 6. This means that we can take a biophysically realistic model of hemodynamic responses (as in Figure 6) and estimate its parameters from measured data using an observation model that is parameterised in terms of kernels with a direct analytic relation to the original parameters θ of the biophysical system (see [49,52] for details). The first-order kernel is simply the conventional HRF. High-order kernels correspond to high-order HRFs and can be estimated using basis functions as described above. In fact, by choosing basis functions as the partial derivatives of the HRF with respect to the biophysical parameters

$$T(\sigma)_i = \frac{\partial \kappa(\sigma)_1}{\partial \theta_i} \tag{9}$$

one can estimate the biophysical parameters from the GLM because, to a first-order approximation, $\beta_i = \theta_i$.

The critical step that we have described in this paragraph is to start with a causal dynamic model of how responses are generated and construct an observation model, using a GLM, that allows us to estimate the parameters of that model. This is in contrast to the conventional use of the GLM with design matrices that are not informed by a forward model of how data are caused. This approach to modelling brain responses has a much more direct connection with underlying physiology and rests upon an understanding of the underlying system.

6.2 Nonlinear system identification—Once a suitable causal model has been established (e.g. Figure 6), we can estimate second-order kernels. These kernels represent a nonlinear characterisation of the HRF that can model interactions among stimuli in causing responses. One important manifestation of the nonlinear effects, captured by the second-order kernels, is

[3]For simplicity, in Eq(7) we deal with only one experimental input.

a modulation of stimulus-specific responses by preceding stimuli that are close in time. This means that responses at high stimulus presentation rates saturate and, in some instances, show an inverted U-shape behaviour. This behaviour appears to be specific to BOLD effects (as distinct from evoked changes in cerebral blood flow) and may represent a *hemodynamic refractoriness*. This effect has important implications for event-related fMRI, where one may want to present trials in quick succession.

The results of a typical nonlinear analysis are given in Figure 7. The results in the right panel represent the average response, integrated over a 32-second train of stimuli as a function of stimulus onset asynchrony (SOA). These responses are based on the kernel estimates (left hand panels) using data from a voxel in the left posterior temporal region of a subject obtained during the presentation of single words at different rates. The solid line represents the estimated response and shows a clear maximum at just less than one second. The dots are responses based on empirical data from the same experiment. The broken line shows the expected response in the absence of nonlinear effects (i.e. that predicted by setting the second order kernel to zero). It is obvious that nonlinearities become important at around two seconds, leading to an actual diminution of the integrated response at sub-second SOAs. The implication of this sort of result is that the assumptions of the linear convolution models discussed above are violated with sub-second SOAs (see also [53,54])

In summary, we started with models of regionally specific responses, framed in terms of the general linear model, in which responses were modelled as linear mixtures of experimentally controlled explanatory variables. Hierarchical extensions to linear observation models enable random-effects analyses and, in particular, empirical Bayesian approaches. These models obtain a mechanistic utility through the use of forward models that embody causal dynamics. Simple variants of these are the linear convolution models used to construct explanatory variables in conventional analyses of fMRI data. These are a special case of generalised convolution models that are mathematically equivalent to input-state-output systems comprising hidden states. Estimation and inference with these dynamic models tells us something about *how* the response was caused. So far, however, we have restricted all models to processes at the level of a single voxel. The next section retains the same perspective on models, but in the context of distributed responses and functional integration.

IV Models of functional integration

1. Functional and effective connectivity

Imaging neuroscience has firmly established functional specialisation as a principle of human brain organisation. The functional integration of the different specialised areas has proven more difficult to assess. Functional integration is generally inferred from simultaneously measured activity of spatially remote neuronal units and is described by means of two different concepts, functional and effective connectivity. *Functional connectivity* has been defined as the correlation among remote neurophysiological events. However, such correlations can arise in a variety of ways: in multi-unit electrode recordings, for example, they can result from stimulus-locked transients evoked by a common input or reflect stimulus-induced oscillations mediated by synaptic connections [55]. Integration within a distributed system is usually better understood in terms of effective connectivity. *Effective connectivity* refers explicitly to the influence that one neural system exerts over another, either at a synaptic (i.e. synaptic efficacy) or population level. A very useful operationalization by Aertsen & Preißl [56] proposes that "the [electrophysiological] notion of effective connectivity should be understood as the experiment- and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons". This speaks to two important points: (i) Effective connectivity is dynamic, i.e. activity- and context-dependent and (ii) it depends upon a model of the interactions. The models for estimating effective connectivity

from functional neuroimaging data used so far can be divided into those based on (i) linear regression models (e.g. Structural Equation Modelling, [57]) or (ii) nonlinear dynamic causal models (e.g. Dynamic Causal Modelling, [58]).

There is a necessary link between functional integration and multivariate analyses because the latter are required to model interactions among brain regions. Multivariate approaches can be divided into those that are inferential in nature and those that are data-led or exploratory. We will first consider exploratory multivariate approaches that are based on covariance patterns, and thus refer to functional connectivity, and then turn to inferential multivariate models of effective connectivity.

1.1 Functional connectivity: principal and independent components analysis—

Friston et al. [59] introduced voxel-based principal component analysis (PCA) of neuroimaging time-series to characterise distributed brain systems implicated in sensorimotor, perceptual or cognitive processes. These distributed systems are identified with principal components or *eigenimages* that correspond to spatial modes of coherent brain activity. This approach represents one of the simplest multivariate characterisations of functional neuroimaging time series and falls into the class of exploratory analyses. Principal component or eigenimage analysis generally uses singular value decomposition (SVD) to identify a set of orthogonal spatial modes that capture the greatest amount of variance expressed over time. As such the ensuing modes embody the most prominent aspects of the variance-covariance structure of a given time series. Because inter-regional covariance is equivalent to functional connectivity eigenimage analysis was one of the first approaches to address functional integration using neuroimaging data. Subsequently, eigenimage analysis has been elaborated in a number of ways. Notable among these is canonical variate analysis (CVA) and multidimensional scaling [60,61]. CVA was introduced in the context of ManCova (multiple analysis of covariance) and uses the generalised eigenvector solution to maximise the variance that can be explained by some explanatory variables relative to error. CVA can be thought of as an extension of eigenimage analysis that refers explicitly to some explanatory variables and allows for statistical inference. A technique closely related to CVA is Partial Least Squares that has been applied in the context of neuroimaging using behavioural data or time series from a reference voxel as explanatory variables [62,63].

In fMRI, eigenimage analysis [64] is generally used as an exploratory device to characterise coherent brain activity. These variance components may, or may not be, related to experimental design. For example, functional connectivity during a “resting state” condition has been observed in the motor system at very low frequencies [65]. Despite its exploratory power, eigenimage analysis is fundamentally limited for two reasons. Firstly, it offers only a linear decomposition of any set of neurophysiological measurements and second, the particular set of eigenimages or spatial modes obtained is uniquely determined by orthogonality constraints that are biologically implausible. These aspects of PCA confer inherent limitations on the interpretability and usefulness of eigenimage analysis of biological time-series and have motivated the exploration of nonlinear PCA and neural network approaches (e.g. [66]).

As a final approach to characterising functional connectivity, independent component analysis (ICA) should be mentioned here. ICA uses entropy maximisation to find, using iterative schemes, spatial modes or their dynamics that are approximately independent. Statistical independence is a stronger requirement than orthogonality in PCA and involves removing high order correlations among the modes (or dynamics). It was initially introduced as *spatial ICA* [67] in which the independence constraint was applied to the modes (with no constraints on their temporal expression). More recent approaches use, by analogy with magneto- and electrophysiological time-series analysis, *temporal ICA* where the dynamics are enforced to

be independent [68]. This requires an initial dimension reduction (usually using conventional eigenimage analysis).

All these approaches are interesting but not used very often. This is largely because they are exploratory and do not allow one to address mechanistic questions about how the brain works. In other words, demonstrating statistical dependencies among regional brain responses (i.e. demonstrating functional connectivity) does not address how these responses were caused. In other words, analyses of functional connectivity do not incorporate any knowledge about the system structure and the causal mechanisms by which the dynamics results from the interaction between external inputs and system structure. However, exploratory techniques can be useful in situations where very little knowledge exists about the system of interest and hypotheses need to be generated by a data-led approach. In most cases, models that embody specific and neurobiologically constrained ideas about how observations were caused, i.e. models of effective connectivity, are more powerful. Generally speaking, models that provide explicit descriptions of causal structure-function relationships within systems are not only required in the context of neuroimaging, but are necessary for understanding the mechanisms underlying any complex system, whether in the domain of biology, physics, or sociology [69].

2. Dynamic causal modelling

This section is about modelling interactions among neuronal populations, at a cortical level, using neuroimaging time series. The aim of these dynamic causal models (DCMs, [58]) is to estimate, and make inferences about, the coupling among brain areas and how that coupling is influenced by changes in experimental context (e.g. time, learning or cognitive set). The basic idea is to construct a reasonably realistic neuronal model of interacting cortical regions or nodes. This model is then supplemented with a forward model of how neuronal or synaptic activity translates into a measured response (see previous section). This enables the parameters of the neuronal model (i.e. effective connectivity) to be estimated from observed data.

Intuitively, this approach regards an experiment as a designed perturbation of neuronal dynamics that are promulgated and distributed throughout a system of coupled anatomical nodes to change region-specific neuronal activity. These changes engender, through a measurement-specific forward model, responses that are used to identify the architecture and time constants of the system at a neuronal level. This represents a departure from conventional approaches (e.g. Structural Equation Modelling and multivariate autoregressive models; [57, 70,71]), in which the inputs to the system are treated as unknown and stochastic and one assumes that the observed responses are driven by intrinsic noise (i.e. innovations). In contradistinction, dynamic causal models assume the responses are driven by experimentally controlled changes in inputs. An important conceptual aspect of dynamic causal models pertains to how the experimental inputs enter the model and cause neuronal responses. Experimental variables can influence the dynamics of the system in one of two ways. First, they can elicit responses through direct influences on specific anatomical nodes (driving inputs). This would be appropriate, for example, in modelling evoked responses in early sensory cortices. The second class of input exerts its effect vicariously, through a modulation of the coupling among nodes (modulatory inputs). These sorts of experimental variables would normally be more enduring; for example, attention to a particular attribute in the stimulus material or the maintenance of some cognitive set. These distinctions are seen most clearly in relation to particular forms of causal models used for estimation, for example the bilinear approximation

$$\begin{aligned}\dot{x} &= f(x,u)\\ &= \left(A+\sum_j u_j B^{(j)}\right)x+Cu\\ y &= g(x)+\varepsilon\\ \theta &= \left\{A, B^{(1)},\ldots,B^{(n)}, C\right\}\\ A &= \frac{\partial \dot{x}}{\partial x} \quad B^{(j)}=\frac{\partial^2 \dot{x}}{\partial u_j \partial x} \quad C=\frac{\partial \dot{x}}{\partial u}\end{aligned} \tag{10}$$

This is an approximation to any model of how changes in neuronal activity in one region x_i are caused by activity in the other regions. Here, effective connectivity is the influence that one neuronal system exerts over another in terms of inducing a response $\partial\dot{x}/\partial x$. The strength of the effective connectivity among the regions in the absence of modulatory inputs is represented by the matrix A. For each input u_j, the matrix $B^{(j)}$ is effectively the change in coupling induced by this input. That means, $B^{(j)}$ encodes the input-sensitive changes in A or, equivalently, the modulation of effective connectivity by the j-th input. Because $B^{(j)}$ is a second-order derivative it is referred to as *bilinear*. Finally, the matrix C embodies the direct (driving) influences of inputs on neuronal activity. The parameters $\theta = \{A, B^{(1)},\ldots,B^{(n)}, C\}$ are the connectivity or coupling matrices that we wish to identify and define the functional architecture and interactions among brain regions at a neuronal level. Finally, the output function $g(x)$ embodies a forward (e.g. hemodynamic) model, linking neuronal activity to measured responses in each region (e.g. for fMRI the model in Figure 6). Note that the inputs u, states x, and observed output y are all functions of time t, whereas the parameters are time-invariant.

Because Eq. 10 has exactly the same form as Eq. 7, we can express it as a GLM and estimate the parameters using EM (see [58]). Generally, estimation in the context of highly parameterised models like DCMs requires constraints in the form of priors. These priors enable conditional inference about the connectivity estimates. The sorts of questions that can be addressed with DCMs are now illustrated by looking at how attentional modulation might be mediated in sensory processing hierarchies in the brain.

2.1 DCM and attentional modulation—It has been established that the superior posterior parietal cortex (SPC) exerts a modulatory role on V5 responses using Volterra-based regression models [72] and that the inferior frontal gyrus (IFG) exerts a similar influence on SPC using structural equation modelling [70]. The example here shows that DCM leads to the same conclusions but starting from a completely different construct. The experimental paradigm and data acquisition parameters are described in the legend to Figure 8. The regions whose time series entered the DCM were based on maxima from conventional SPMs testing for the effects of photic stimulation, motion and attention. Regional time courses were taken as the first eigenvariate of 8mm spherical volumes of interest centred on the local maxima in the SPMs (see [58] for details). The inputs, in this example, comprise one sensory perturbation and two contextual inputs. The sensory input was simply the presence of photic stimulation and the first contextual input was presence of motion in the visual field. The second contextual input, encoding attentional set, was unity during attention to speed changes and zero otherwise. The outputs corresponded to the four regional eigenvariates in Figure 8 (see plots on the right). The intrinsic connections were constrained to conform to a hierarchical pattern in which each area was reciprocally connected to its supraordinate area. Photic stimulation entered at, and only at, V1. The effect of motion in the visual field was modelled as a bilinear modulation of the V1 to V5 connectivity and attention was allowed to modulate the backward connections from IFG and SPC.

The results of the DCM are shown in Figure 8 (right panel). Of primary interest here is the modulatory effect of attention that is expressed in terms of the bilinear coupling parameters for this input. As expected, we can be highly confident that attention modulates the backward connections from IFG to SPC and from SPC to V5. Indeed, the influences of IFG on SPC were negligible in the absence of attention. It is important to note that, in this model, the only way that attentional manipulation could affect brain responses was through this bilinear effect on connection strengths. This change is, presumably, instantiated by instructional set at the beginning of each epoch.

The second thing that this analysis illustrates is how functional segregation is modelled in DCM. Here one can regard V1 as 'segregating' motion from other visual information and distributing it to the motion-sensitive area V5. This segregation is modelled as a bilinear 'enabling' of V1 to V5 connections when, and only when, motion is present. Note that in the absence of motion the intrinsic V1 to V5 connection was trivially small (in fact the estimate was −0.04). The key advantage of entering motion through a bilinear effect, as opposed to a direct effect on V5, is that we can finesse the inference that V5 shows motion-selective responses with the assertion that these responses are mediated by afferents from V1. The two bilinear effects described above represent two important aspects of functional integration that DCM is able characterise.

2.2 Structural equation modelling—The central idea behind dynamic causal modelling (DCM) is to treat the brain as a deterministic nonlinear dynamic system that receives external inputs and produces outputs. Effective connectivity is parameterised in terms of coupling among unobserved brain states (e.g. neuronal activity in different regions). The objective is to estimate these parameters by perturbing the system and measuring the response. This is in contradistinction to established methods for estimating effective connectivity from neurophysiological time series, which include structural equation modelling (SEM) and models based on multivariate auto-regressive processes. In these models, there is no designed perturbation, and the inputs are treated as unknown and stochastic. Furthermore, the inputs are assumed to express themselves instantaneously[4] such that, at each point of observation, the system is at equilibrium and the change in states will be zero. If we reformulate Eq. 10 under this assumption, treat the inputs as random innovations and omit bilinear effects, we obtain the regression equation used in SEM:

$$\begin{aligned} \dot{x} &= Ax + Cu = 0 \Rightarrow \\ x &= -A^{-1}Cu \end{aligned} \tag{11}$$

The key point here is that A is estimated by assuming u is some random innovation with known covariance. This is suboptimal for designed experiments, where u represents carefully structured experimental inputs, because we are throwing away information. SEM and multivariate autoregressive models are certainly useful for establishing dependencies among observed regional responses, but not optimal for designed perturbations or experiments.

In this section we have covered multivariate techniques ranging from eigenimage analysis that does not have an explicit forward or causal model to DCM that does. The bilinear approximation to any DCM has been illustrated through its use with fMRI to study attentional modulation. Although the bilinear approximation described above is useful, it is possible to model effective connectivity among neuronal subpopulations even more directly. We now move on to a DCM that embraces a number of neurobiological facts and takes us much closer

[4]In principle, one could extend SEM to incorporate inputs with temporal lag, but this is rarely done in practice.

to a mechanistic understanding of how brain responses are generated. This example uses responses measured with EEG.

3. Dynamic causal modelling with neural mass models

Event-related potentials (ERPs) have been used for decades as electrophysiological correlates of perceptual and cognitive operations. However, the exact neurobiological mechanisms underlying their generation are largely unknown. In this section we introduce a biologically plausible model to understand event-related responses. The example used in this section shows that changes in connectivity are sufficient to explain certain ERP components. Specifically we will look at the P300, a late component associated with rare or unexpected events. If the unexpected nature of rare stimuli depends on learning which stimuli are frequent, then the P300 must be due to plastic changes in connectivity that mediate perceptual learning. We conclude by showing that recent advances in the modelling of evoked responses now afford measures of connectivity among cortical sources that can be used to quantify the effects of perceptual learning.

3.1 Hierarchical neural mass models—David et al. [73,74] have developed a model of event-related potentials that rests on the connectivity rules summarised by Felleman & Van Essen [75] to assemble a network of coupled cortical sources. These rules are based on distinguishing connections with respect to their laminar patterns of origin and termination. In short, by dividing six-layered iscortical areas into the granular layer (layer 4), supra-granular layers (layers 1-3) and infra-granular layers (layers 5-6), different types of connections can be defined as follows. Bottom-up or forward connections originate in agranular layers and terminate in layer 4. Top-down or backward connections originate and terminate in agranular layers. Lateral connections originate in agranular layers and target all layers. These long-range or extrinsic cortico-cortical connections are excitatory (using glutamate as neurotransmitter) and arise from pyramidal cells.

Each region or source is modelled using a neural mass model described in [73], based on the model of Jansen & Rit [76]. This model emulates the activity of a cortical area using three neuronal subpopulations, assigned to granular and agranular layers. A population of excitatory pyramidal (output) cells receives inputs from inhibitory and excitatory populations of interneurons, via intrinsic connections (intrinsic connections are confined to the cortical sheet). Within this model, excitatory interneurons can be regarded as spiny stellate cells found predominantly in layer 4 and in receipt of forward connections. Excitatory pyramidal cells and inhibitory interneurons will be considered to occupy agranular layers and receive backward and lateral inputs (see Figure 9).

To model event-related responses, the network receives inputs via input connections. These connections are exactly the same as forward connections and deliver inputs u to the spiny stellate cells in layer 4. In the present context, inputs u model subcortical auditory inputs. The parameter vector C controls the influence of the input on each source. The parameter matrices A^F, A^B, A^L encode forward, backward and lateral connections respectively. The DCM here is specified in terms of the state equations shown in Figure 9 and a linear output equation

$$\begin{aligned} \dot{x} &= f(x, u, \theta) \\ y &= Lx_0 + \varepsilon \end{aligned} \tag{12}$$

where x_0 represents the transmembrane potential of pyramidal cells and L is a lead field matrix coupling electrical sources to the EEG channels. This should be compared to the DCM above for hemodynamics. Here the equations governing the evolution of neuronal states are much more complicated and realistic, as opposed to the bilinear approximation in Eq. 10. Conversely,

the output equation is a simple linearity, as opposed to the nonlinear observation equation used for fMRI. As an example, the state equation for the inhibitory subpopulation is [5]

$$\begin{aligned}\dot{x}_7 &= x_8 \\ \dot{x}_8 &= \frac{H_e}{\tau_e}\left(\left(A^B + A^L + \gamma_3 I\right) S\left(x_0\right)\right) - \frac{2x_8}{\tau_e} - \frac{x_7}{\tau_e^2}\end{aligned} \tag{13}$$

Within each subpopulation, the evolution of neuronal states rests on two operators. The first transforms the average density of pre-synaptic inputs into the average postsynaptic membrane potential. This is modelled by a linear transformation with excitatory (e) and inhibitory (i) kernels parameterised by $H_{e,i}$ and $\tau_{e,i}$. $H_{e,i}$ control the maximum post-synaptic potential and $\tau_{e,i}$ represent a lumped rate constant. The second operator S transforms the average potential of each subpopulation into an average firing rate. This is assumed to be instantaneous and is a sigmoid function. Interactions among the subpopulations depend on constants $\gamma_{1,2,3,4}$, which control the strength of intrinsic connections and reflect the total number of synapses expressed by each subpopulation. In Eq. 13, the top line expresses the rate of change of voltage as a function of current. The second line specifies how current changes as a function of voltage, current and presynaptic input from extrinsic and intrinsic sources. Having specified the DCM one can estimate the coupling parameters from empirical data using EM as described above.

3.2 Perceptual learning and the P300—The example shown in Figure 10 is an attempt to model the P300 in terms of changes in backward and lateral connections among cortical sources. In this example, two EEG time series (*i.e*. the averages over two subsets of channels; see circles in Fig. 9) were modelled with three cortical sources[6]. Using this generative or forward model we estimated differences in the strength of these connections for rare and frequent stimuli. As expected, we could account for detailed differences in the ERPs (the P300) by changes in connectivity (see figure legend for details). Interestingly these differences were expressed selectively in the lateral connections. If this model is a sufficient approximation to the real sources, these changes are a non-invasive measure of plasticity, mediating perceptual learning, in the human brain.

Conclusion

In this article we have reviewed some key models of neuroimaging data used to address questions of functional specialisation and integration. In the order that these models were discussed, they embodied increasing amounts of information about how signals measured by neuroimaging techniques like fMRI are generated, both in terms of biophysics and the underlying neuronal interactions. We have seen how hierarchical linear observation models can encode experimentally designed effects. General linear models based on convolution models imply an underlying dynamic input-state-output system. The form of these systems can be used to constrain convolution models and explore some of their simpler nonlinear properties. By creating observation models based on explicit forward models of neuronal interactions, one can now start to model and assess interactions among distributed cortical areas and make inferences about coupling at the neuronal level.

During the next years, the dynamic causal models introduced above are likely to become more and more neurobiologically realistic (see [77]). As shown above, there are already plausible models of neuronal ensembles to estimate network parameters of evoked responses in EEG

[5]For simplicity, propagation delays on the extrinsic connections have been omitted here and in Figure 9.

[6]Note that averaged time series are used for reasons of computational expediency. As the data from all network nodes are concatenated into one single data vector, the resulting covariance matrices would become computationally intractable for long EEG time series with a lot of individual trials.

[73,74]. Other modelling approaches, which are based on mean field approaches and are currently under development, even distinguish between different types of receptors [78]. In the nearer future, these developments are likely to encompass fMRI signals, enabling the conjoint modelling, or fusion, of different neuroimaging modalities.

Acknowledgments

This work was supported by the Wellcome Trust. We would like to thank our colleagues at the Wellcome Department of Imaging Neuroscience for helpful discussions and suggestions.

References

1. Toga, AW.; Mazziotta, JC. Brain Mapping – The Methods. Academic Press; New York: 2002.
2. Ashburner J, Friston KJ. Voxel-based morphometry - the methods. NeuroImage 2000;11:805–821. [PubMed: 10860804]
3. Friston K. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? Annu Rev Neurosci 2002;25:221–250. [PubMed: 12052909]
4. Staum M. Physiognomy and phrenology at the Paris Athénée. J Hist Ideas 1995;56:443–462. [PubMed: 11639800]
5. Phillips CG, Zeki S, Barlow HB. Localisation of function in the cerebral cortex: Past, present and future. Brain 1984;107:327–361. [PubMed: 6421455]
6. Goltz, F. In: MacCormac, W., editor. Transactions of the 7th international medical congress; JW Kolkmann; London. 1881. p. 218-228.
7. Absher JR, Benson DF. Disconnection syndromes: an overview of Geschwind's contributions. Neurology 1993;43:862–867. [PubMed: 8492937]
8. Young MP. The organization of neural systems in the primate cerebral cortex. Proc R Soc Lond B Biol Sci 1993;252:13–18.
9. Passingham RE, Stephan KE, Kotter R. The anatomical basis of functional localization in the cortex. Nat Rev Neurosci 2002;3:606–616. [PubMed: 12154362]
10. Zeki, S. The motion pathways of the visual cortex. In: Blakemore, C., editor. Vision: coding and efficiency. Cambridge University Press; Cambridge: 1990. p. 321-345.
11. Shipp S, Zeki S. The Organization of Connections between Areas V5 and V2 in Macaque Monkey Visual Cortex. Eur J Neurosci 1999;1:333–354. [PubMed: 12106143]
12. Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. Comparing functional (PET) images: the assessment of significant change. J Cereb Blood Flow Metab 1991;11:690–699. [PubMed: 2050758]
13. Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for rCBF activation studies in human brain. J Cereb Blood Flow Metab 1992;12:900–918. [PubMed: 1400644]
14. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach of determining significant signals in images of cerebral activation. Hum Brain Mapp 1996;4:58–73. [PubMed: 20408186]
15. Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ. Statistical Parametric Maps in functional imaging: A general linear approach. Hum Brain Mapp 1995;2:189–210.
16. Friston KJ, Penny W. Posterior probability maps and SPMs. NeuroImage 2003;19:1240–1249. [PubMed: 12880849]
17. Berry DA, Hochberg Y. Bayesian perspectives on multiple comparisons. J Statistical Planning Inference 1999;82:215–227.
18. Kiebel, S.; Holmes, AP. The General Linear Model. In: Frackowiack, R., editor. Human Brain Function. Elsevier; New York: 2004. p. 725-760.
19. Büchel C, Wise RJS, Mummery CJ, Poline JB, Friston KJ. Nonlinear regression in parametric activation studies. NeuroImage 1996;4:60–66. [PubMed: 9345497]
20. Friston KJ, Frith CD, Turner R, Frackowiak RSJ. Characterising evoked hemodynamics with fMRI. NeuroImage 1995;2:157–165. [PubMed: 9343598]

21. Friston KJ, Jezzard PJ, Turner R. Analysis of functional MRI time-series. Hum Brain Mapp 1994;1:153–171.
22. Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited - again. NeuroImage 1995;2:173–181. [PubMed: 9343600]
23. Bullmore ET, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter TA, Brammer M. Colored Noise and Computational Inference in Neurophysiological (fMRI) Time Series Analysis: Resampling Methods in Time and Wavelet Domains. Hum Brain Mapp 2001;12:61–78. [PubMed: 11169871]
24. Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J. Classical and Bayesian inference in neuroimaging: applications. NeuroImage 2002;16:484–512. [PubMed: 12030833]
25. Friston KJ, Josephs O, Zarahn E, Holmes AP, Rouquette S, Poline J. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. NeuroImage 2000;12:196–208. [PubMed: 10913325]
26. Kiebel S, Friston KJ. Statistical parametric mapping for event-related potentials: I. Generic considerations. NeuroImage 2004;22:492–502. [PubMed: 15193578]
27. Kiebel S, Friston KJ. Statistical parametric mapping for event-related potentials (II): a hierarchical temporal model. NeuroImage 2004;22:503–520. [PubMed: 15193579]
28. Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME. Positron emission tomographic studies of the processing of single words. J Cogn Neurosci 1989;1:153–170.
29. Lueck CJ, Zeki S, Friston KJ, Deiber MP, Cope NO, et al. The colour centre in the cerebral cortex of man. Nature 1989;340:386–389. [PubMed: 2787893]
30. Price CJ, Friston KJ. Cognitive Conjunction: A new approach to brain activation experiments. NeuroImage 1997;5:261–270. [PubMed: 9345555]
31. Grafton S, Mazziotta J, Presty S, Friston KJ, Frackowiak RSJ, Phelps M. Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. J Neurosci 1992;12:2542–2548. [PubMed: 1613546]
32. Price CJ, Wise RJS, Ramsay S, Friston KJ, Howard D, et al. Regional response differences within the human auditory cortex when listening to words. Neurosci Lett 1992;146:179–182. [PubMed: 1491785]
33. Chawla D, Phillips J, Büchel C, Edwards R, Friston KJ. Speed-dependent motion-sensitive responses in V5: an fMRI study. NeuroImage 1998;7:86–96. [PubMed: 9558641]
34. Friston KJ, Frith CD, Passingham RE, Liddle PF, Frackowiak RSJ. Motor practice and neurophysiological adaptation in the cerebellum: A positron tomography study. Proc Roy Soc Lond Series B 1992;248:223–228.
35. Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RSJ, Dolan RJ. The trouble with cognitive subtraction. NeuroImage 1996;4:97–104. [PubMed: 9345501]
36. Genovese CR, Lazar NR, Nichols T. Thesholding of statistical in functional neuroimaging using the false discovery rate. NeuroImage 2002;15:870–878. [PubMed: 11906227]
37. Friston KJ, Stephan KE, Lund TE, Morcom A, Kiebel S. Mixed-effects and fMRI studies. NeuroImage 2005;24:244–252. [PubMed: 15588616]
38. Dempster AP, Laird NM, Rubin DR. Maximum likelihood from incomplete data via the EM algorithm. J Roy Stat Soc Series B 1977;39:1–38.
39. Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G, Ashburner J. Classical and Bayesian inference in neuroimaging: theory. NeuroImage 2002;16:465–483. [PubMed: 12030832]
40. Phillips C, Mattout J, Rugg MD, Maquet P, Friston KJ. Parametric empirical Bayes solution of the source reconstruction problem in EEG. NeuroImage 2005;24:997–1011. [PubMed: 15670677]
41. Boynton GM, Engel SA, Glover GH, Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 1996;16:4207–4221. [PubMed: 8753882]
42. Josephs O, Turner R, Friston KJ. Event-related fMRI. Hum Brain Mapp 1997;5:243–248. [PubMed: 20408223]
43. Lange N, Zeger SL. Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging. J Roy Stat Soc C 46:1–29.

44. Handwerker DA, Ollinger JM, D'Esposito M. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 2004;21:1639–1651. [PubMed: 15050587]
45. Marrelec G, Benali H, Ciuciu P, Pelegrini-Issac M, Poline JB. Robust Bayesian estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. Hum Brain Mapp 19:1–17. [PubMed: 12731100]
46. Logothetis NK, Wandell BA. Interpreting the BOLD signal. Annu Rev Physiol 2004;66:735–769. [PubMed: 14977420]
47. Stephan KE, Harrison LM, Penny WD, Friston KJ. Biophysical models of fMRI responses. Curr Opin Neurobiol 2004;14:629–635. [PubMed: 15464897]
48. Buxton RB, Frank LR. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. J Cereb Blood Flow Metab 1997;17:64–72. [PubMed: 8978388]
49. Friston KJ, Mechelli A, Turner R, Price CJ. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 2000;12:466–477. [PubMed: 10988040]
50. Fliess M, Lamnabhi M, Lamnabhi-Lagarrigue F. An algebraic approach to nonlinear functional expansions. IEEE Trans Circuits Syst 1983;30:554–570.
51. Bendat, JS. Nonlinear System Analysis and Identification from Random Data. John Wiley and Sons; New York: 1990.
52. Friston KJ. Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage 2002;16:513–530. [PubMed: 12030834]
53. Buckner R, Bandettini P, O'Craven K, Savoy R, Petersen S, Raichle M, Rosen B. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. Proc Natl Acad Sci USA 1996;93:14878–14883. [PubMed: 8962149]
54. Burock MA, Buckner RL, Woldorff MG, Rosen BR, Dale AM. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. NeuroReport 9:3735–3739. [PubMed: 9858388]
55. Gerstein GL, Perkel DH. Simultaneously recorded trains of action potentials: Analysis and functional interpretation. Science 1969;164:828–830. [PubMed: 5767782]
56. Aertsen, A.; Preißl, H. Dynamics of activity and connectivity in physiological neuronal networks. In: Schuster, HG., editor. Nonlinear dynamics and neuronal networks. VCH Publishers Inc.; New York: 1991. p. 281-302.
57. McIntosh AR, Gonzalez-Lima F. Structural equation modelling and its application to network analysis in functional brain imaging. Hum Brain Mapp 1994;2:2–22.
58. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. NeuroImage 2003;19:1273–1302. [PubMed: 12948688]
59. Friston KJ, Frith CD, Liddle PF, Frackowiak RSJ. Functional Connectivity: The principal component analysis of large data sets. J Cereb Blood Flow Metab 1993;13:5–14. [PubMed: 8417010]
60. Friston KJ, Poline JB, Holmes AP, Frith CD, Frackowiak RSJ. A multivariate analysis of PET activation studies. Hum Brain Mapp 1996;4:140–151. [PubMed: 20408193]
61. Friston KJ, Frith CD, Fletcher P, Liddle PF, Frackowiak RSJ. Functional topography: multidimensional scaling and functional connectivity in the brain. Cereb Cortex 1996;6:156–164. [PubMed: 8670646]
62. McIntosh AR, Bookstein FL, Haxby JV, Grady CL. Spatial pattern analysis of functional brain images using partial least squares. NeuroImage 1996;3:143–157. [PubMed: 9345485]
63. McIntosh AR, Chau WK, Protzner AB. Spatiotemporal analysis of event-related fMRI data using partial least squares. NeuroImage 2004;23:764–775. [PubMed: 15488426]
64. Sychra JJ, Bandettini PA, Bhattacharya N, Lin Q. Synthetic images by subspace transforms. I. Principal component images and related filters. Med Physics 1994;21:193–201.
65. Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Res Med 1995;34:537–541.
66. Friston KJ, Phillips J, Chawla D, Büchel C. Nonlinear PCA: characterizing interactions between modes of brain activity. Philos Trans R Soc Lond B Biol Sci 355:135–146. [PubMed: 10703049]

67. McKeown MJ, Jung TP, Makeig S, Brown G, Kinderman S, Lee TW, Sejnowski T. Spatially independent activity patterns in functional MRI data during the Stroop colour naming task. Proc Natl Acad Sci 1998;95:803–810. [PubMed: 9448244]

68. McKeown MJ, Hansen LK, Sejnowski T. Independent component analysis of functional MRI: what is signal and what is noise? Curr Opin Neurobiol 2003;13:620–629. [PubMed: 14630228]

69. Stephan KE. On the role of general systems theory for functional neuroimaging. J Anat 2004;205:443–470. [PubMed: 15610393]

70. Büchel C, Friston KJ. Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. Cereb Cortex 1997;7:768–778. [PubMed: 9408041]

71. Harrison LM, Penny W, Friston KJ. Multivariate autoregressive modelling of fMRI time series. NeuroImage 2003;19:1477–1491. [PubMed: 12948704]

72. Friston KJ, Büchel C. Attentional modulation of effective connectivity from V2 to V5/MT in humans. Proc Natl Acad Sci USA 2000;97:7591–7596. 2000. [PubMed: 10861020]

73. David O, Friston KJ. A neural mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 2003;20:1743–1755. [PubMed: 14642484]

74. David, O.; Harrison, L.; Kilner, J.; Penny, W.; Friston, KJ. Studying effective connectivity with a neural mass model of evoked MEG/EEG responses. In: Halgren, E.; Ahlfors, S.; Hämäläinen, M.; Cohen, D., editors. Proceedings of the 14th international conference on biomagnetism BIOMAG; Boston, Ma., USA. 2004. p. 135-138.

75. Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1991;1:1–47. [PubMed: 1822724]

76. Jansen BH, Rit VG. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol Cybern 1995;73:357–366. [PubMed: 7578475]

77. Horwitz B, Friston KJ, Taylor JG. Neural modelling and functional brain imaging: an overview. Neural Netw 2001;13:829–846. [PubMed: 11156195]

78. Harrison LM, David O, Friston KJ. Dynamic mean fields and ERP generation. Philos Trans R Soc Lond B Biol Sci. 2005 in press.

79. Dale A, Buckner R. Selective averaging of rapidly presented individual trials using fMRI. Hum Brain Mapp 1997;5:329–340. [PubMed: 20408237]

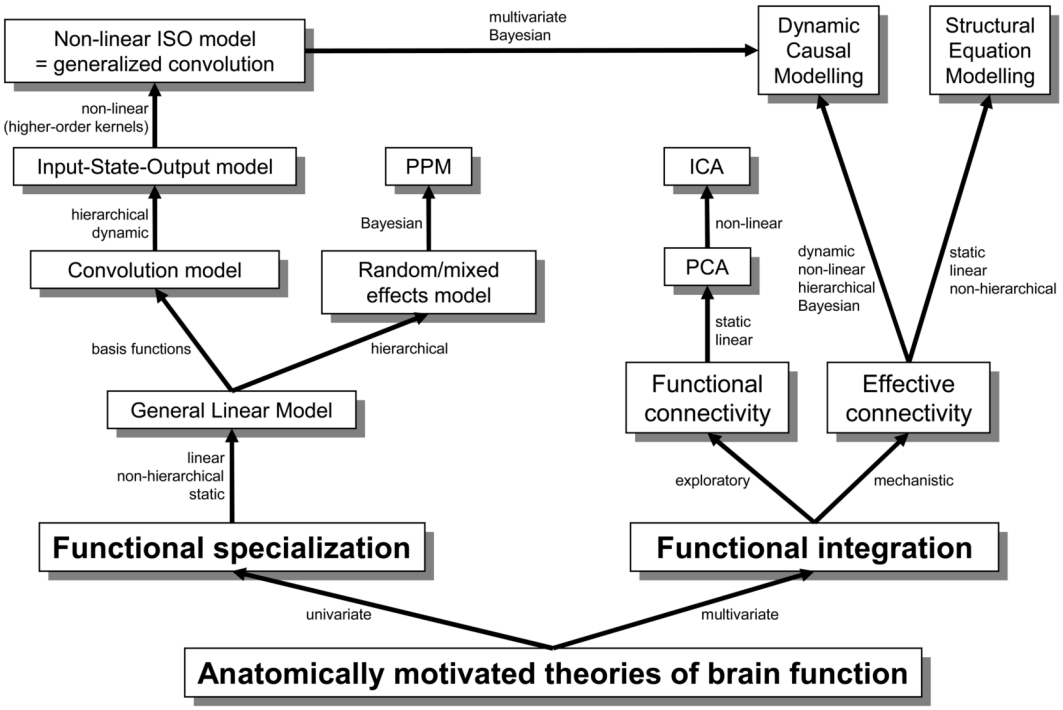

Figure 1.

This diagram shows how the models discussed in this article are related to each other. The exposition of the various models in this article follows the hierarchical relations shown in this figure. Abbreviations: ISO = input-state-output, ICA = independent components analysis, PCA = principal components analysis, PPM = posterior probability map.

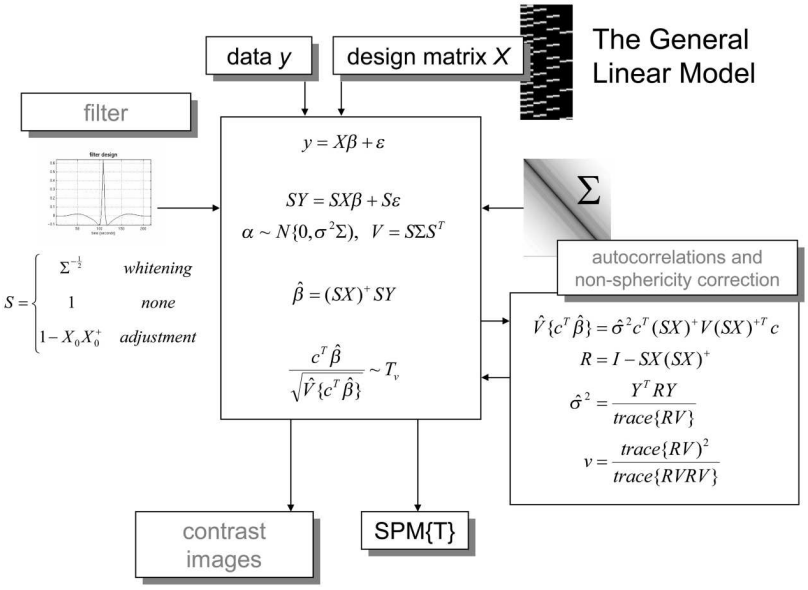

Figure 2.

The general linear model is an equation expressing the response variable *y* in terms of a linear combination of explanatory variables, represented by the columns of the design matrix *X*, and an error term *ε* with assumed or known autocorrelation Σ [18]. In fMRI, the data can be filtered with a convolution or residual forming matrix (or a combination) *S*, leading to a generalised linear model that includes [intrinsic] serial correlations and applied [extrinsic] filtering. Different choices of *S* correspond to different estimation schemes as indicated on the upper left. The parameter estimates obtain in a least squares sense using the pseudoinverse (denoted by +) of the filtered design matrix. An effect of interest is specified by a vector of contrast weights *c* that give a weighted sum or compound of parameter estimates, referred to as a *contrast*. The T statistic is simply this contrast divided by the standard error (i.e. square root of its estimated variance). The ensuing T statistic is distributed with *v* degrees of freedom. The equations for estimating the variance of the contrast and the degrees of freedom are provided in the right-hand panel and accommodate the non-sphericity implied by Σ.

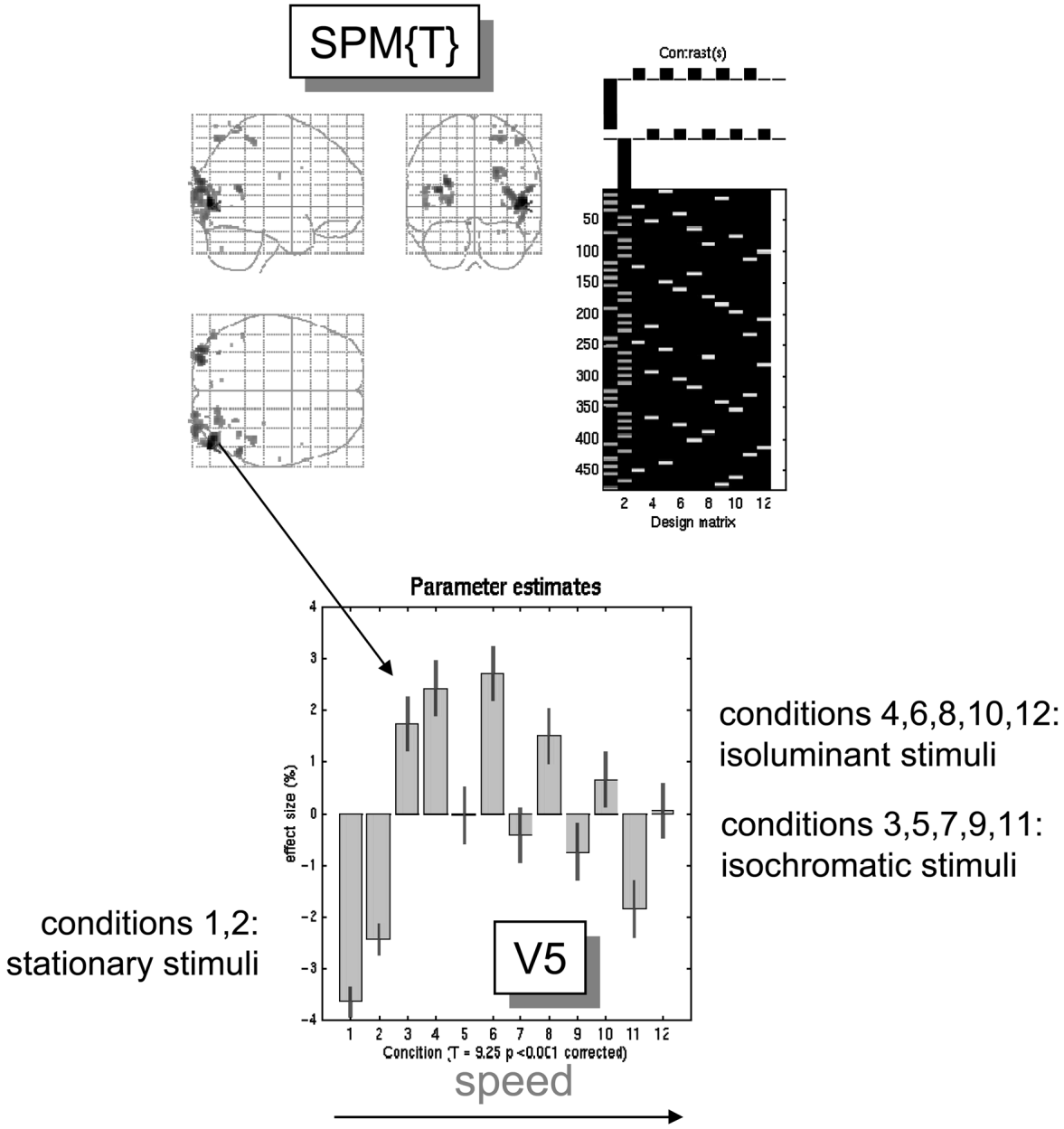

Figure 3.

Example of an experiment whose design has both factorial and parametric properties [33]. Top right: Design matrix: This is an image representation of the design matrix. Contrasts: These are the vectors of contrast weights defining the linear compounds of parameters tested. The contrast weights are displayed over the column of the design matrix encoding the effects tested. The design matrix here includes condition-specific effects (boxcars convolved with a hemodynamic response function). Odd columns correspond to stimuli shown under isochromatic conditions and even columns model responses to isoluminant stimuli. The first two columns are for stationary stimuli and the remaining columns are for stimuli of increasing speed. The final column is a constant term. Top left: SPM{T}: This is a maximum intensity projection of the SPM{T} conforming to standard MNI (Montreal Neurological Institute) space based on the Talairach & Tournoux (1988) system. The T values here are the minimum T

values from both contrasts, thresholded at $p = 0.001$ uncorrected. The most significant conjunction is seen in left V5. Lower panel: Plot of the condition-specific parameter estimates for this voxel. The T value was 9.25 ($p<0.001$ adjusted according to RFT).

UKPMC Funders Group Author Manuscript

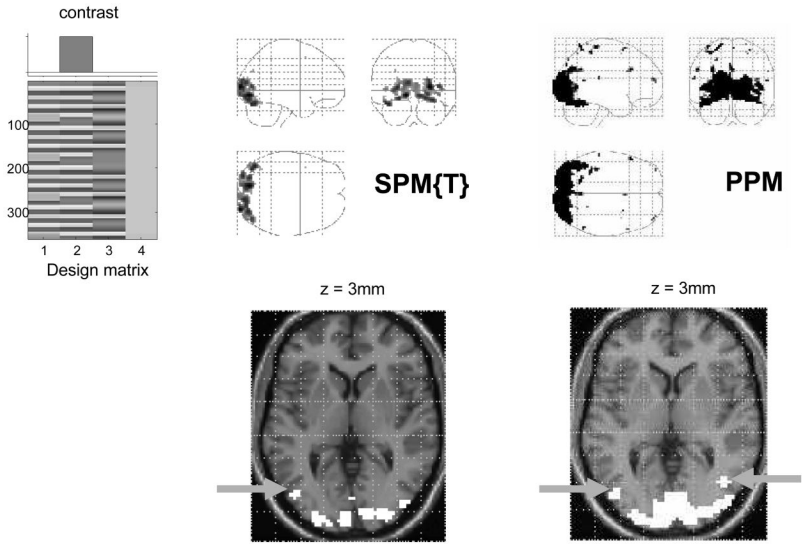

Figure 4.

SPM and PPM for an fMRI study of attention to visual motion [70]. The display format in the lower panel uses an axial slice through extrastriate regions but the thresholds are the same as employed in the maximum intensity projections (upper panels). Upper right: PPM showing all voxels that exceeded a 90% chance of an activation of 0.7% global mean signal (arbitrary units). Upper left: The corresponding SPM using an adjusted threshold at $p < 0.05$. Note the bilateral foci of motion-related responses in the PPM that are not seen in the SPM (grey arrows). As can be imputed from the design matrix (upper left panel), the statistical model of evoked responses comprised boxcar regressors convolved with a canonical hemodynamic response function. The middle column corresponds to the presentation of moving dots and was the stimulus property tested by the contrast.

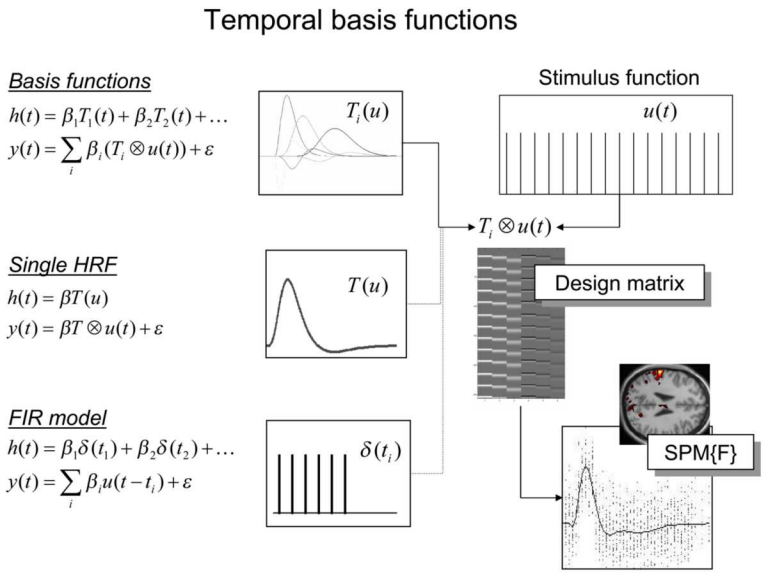

Figure 5.

Temporal basis functions offer useful constraints on the form of the estimated response that retain the flexibility of FIR models and the efficiency of single regressor models (see [20,42, 43] for details). The specification of a GLM that rests on temporal basis functions requires stimulus functions $u(t)$ (top right) that model expected neuronal changes (e.g. boxcars of epoch-related responses or delta functions representing specific events). These stimulus functions are then convolved with a set of basis functions $T_i(t)$ of peri-stimulus time (top middle) that, in some linear combination, model the HRF (top left; see also Eq. 6). The resulting time series enter as regressors into the design matrix (middle right). The basis functions can be as simple as a single canonical HRF (middle), through to a series of delta functions for each time point following the onset of a trial type (middle bottom). The latter case corresponds to a FIR model (left bottom); here, the parameters estimates describe the impulse response function at a finite number of discrete sampling times. Note that selective averaging in event-related fMRI [79] is mathematically equivalent to this limiting case.

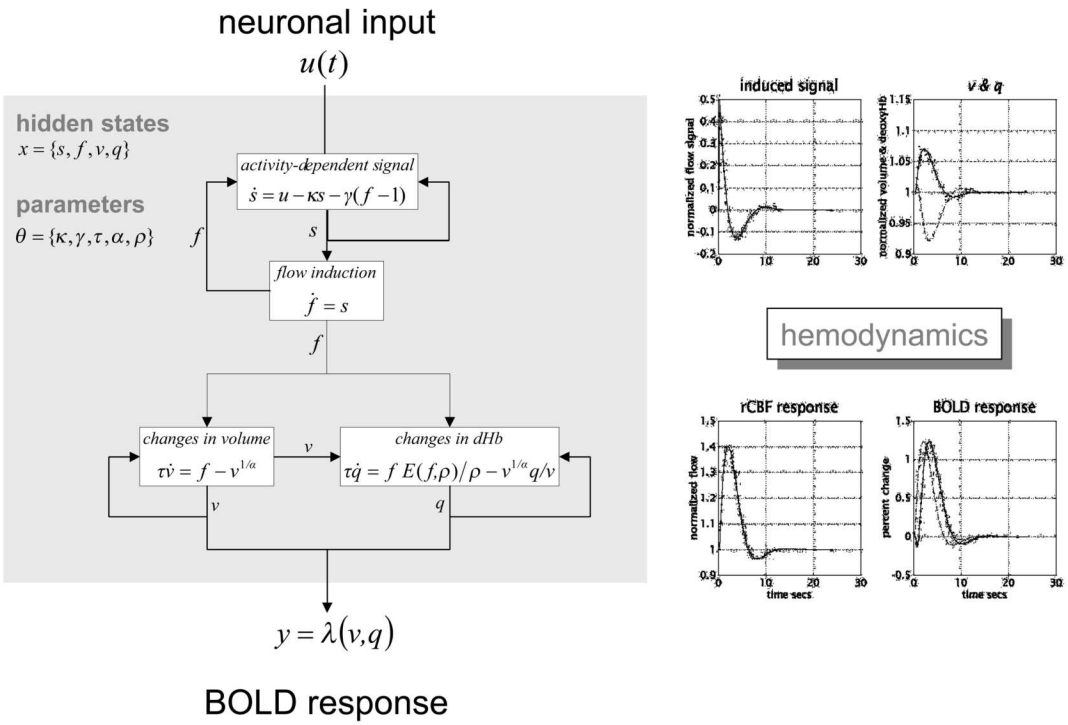

Figure 6.

Right: Hemodynamics elicited by an impulse of neuronal activity as predicted by a dynamical biophysical model (left). A burst of neuronal activity causes an increase in flow-inducing signal that decays with first order kinetics and is down regulated by local flow. This signal increases rCBF, which dilates the venous capillaries, increasing volume (v). Concurrently, venous blood is expelled from the venous pool decreasing deoxyhemoglobin content (q). The resulting fall in deoxyhemoglobin concentration leads to a transient increases in BOLD (blood oxygenation level dependent) signal and a subsequent undershoot. Left: Hemodynamic model on which these simulations were based (see [49,52] for details).

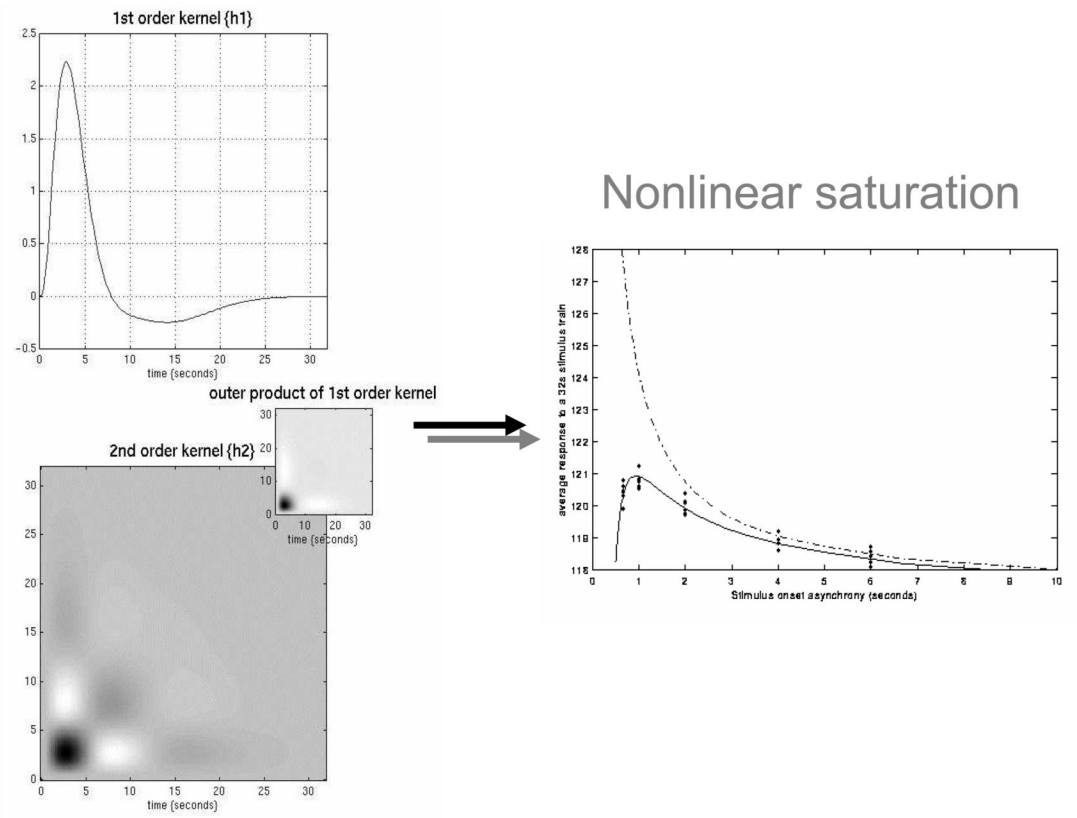

Figure 7.

Left panels: Volterra kernels from a voxel in the left superior temporal gyrus at -56, -28, 12mm. These kernel estimates were based on a single-subject study of aural word presentation at different rates (from 0 to 90 words per minute) using a second order approximation to a Volterra series expansion modelling the observed hemodynamic response to stimulus input (a delta function for each word). These kernels can be thought of as a characterisation of the second order hemodynamic response function. The first order kernel κ_1 (upper panel) represents the (first-order) component usually presented in linear analyses. The second-order kernel (lower panel) is presented in image format. The colour scale is arbitrary; white is positive and black is negative. The insert on the right represents $\kappa_1 \kappa_1^T$, the second-order kernel predicted by a simple model that involved a linear convolution with κ_1 followed by some static nonlinearity. Right panel: Integrated responses over a 32-second stimulus train as a function of SOA. Solid line: Estimates based on the nonlinear convolution model parameterised by the kernels on the left. Broken line: The responses expected in the absence of second-order effects (*i.e.* in a truly linear system). Dots: Empirical averages based on the presentation of actual stimulus trains.

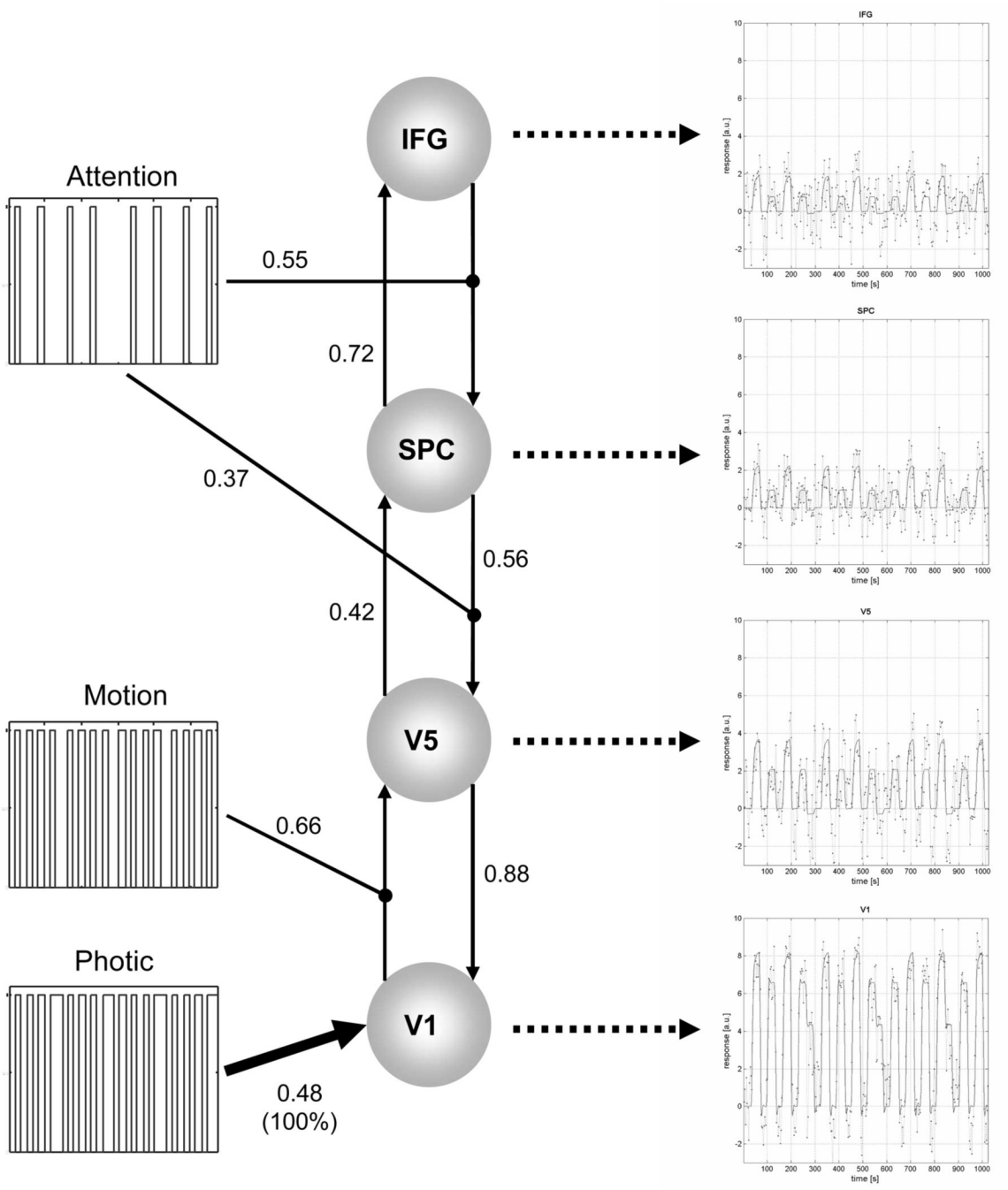

Figure 8.

DCM analysis of a single subject fMRI data set on attention to visual motion. The fMRI data were from a study in which subjects viewed identical stimuli (radially moving dots) under different attentional manipulations of the task (detection of velocity changes) – see [70]. Only those conditional estimates are shown alongside their connections for which there was at least 90% confidence that they exceeded the chosen threshold of 0.17 Hz (corresponding to neural transients with a half life shorter than 4 seconds). The shown values resulted from a re-analysis with the developer version of SPM2 (as of May 2004) and therefore marginally diverge from those reported previously by Friston et al. [58]. The temporal structure of the inputs is shown by box-car plots (left). Note that motion and attention exert bilinear effects: motion modulates

the connection from V1 to the motion-sensitive area V5, whereas attention modulates the backward connections from the inferior frontal gyrus (IFG) to the superior parietal cortex (SPC) and from SPC to V5. Fitted responses based upon the conditional estimates and the adjusted data are shown in the panels connected to the areas by dotted arrows.

UKPMC Funders Group Author Manuscript

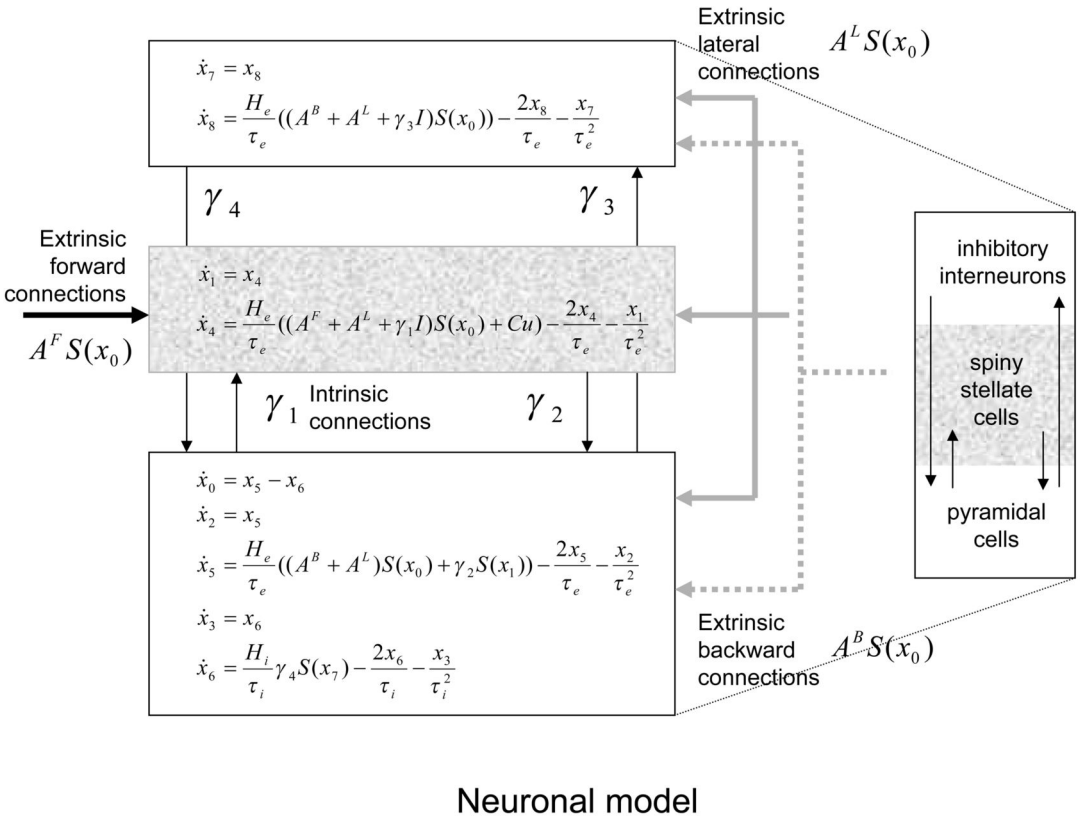

Figure 9.
Schematic of the DCM used to model electrical responses. This schematic shows the state equation describing the dynamics of sources or regions. Each source is modelled with three subpopulations (pyramidal, spiny stellate and inhibitory interneurons); see [73,76] for details. These have been assigned to granular and agranular cortical layers which receive forward and backward connections respectively.

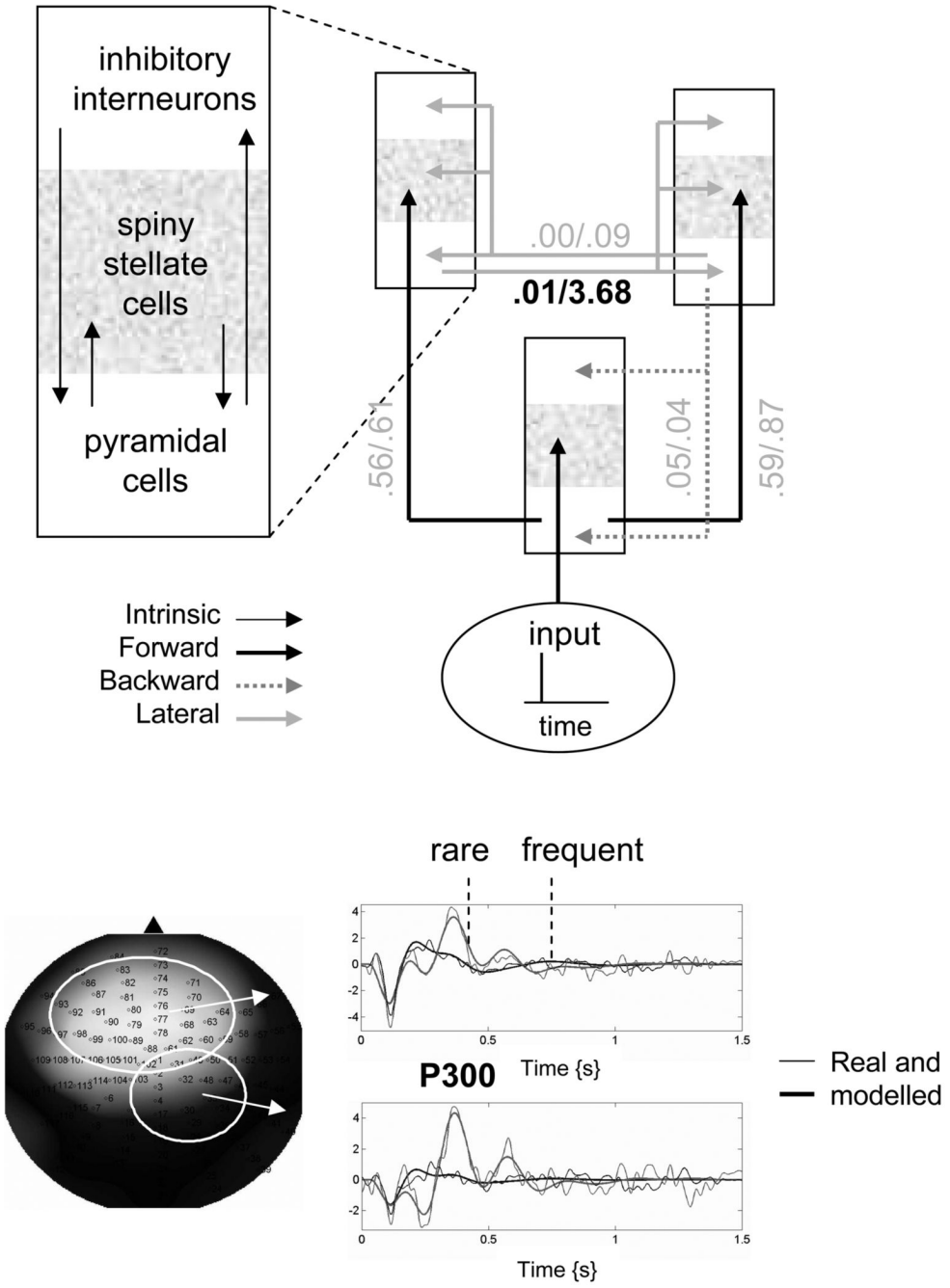

Figure 10.

Summary of a dynamic causal modelling of ERPs elicited during an auditory P300 paradigm, employing rare and frequent pure tones. Upper panel: Schematic showing the architecture of the neuronal model used to explain the empirical data. Sources were coupled with extrinsic cortico-cortical connections following the rules of Felleman and van Essen [75]. The free parameters of this model included intrinsic and extrinsic connection strengths that were adjusted to best explain the observed ERPs. In this example the lead field was also estimated, with no spatial constraints. The parameters were estimated for ERPs recorded during the presentation of rare and frequent tones and are reported beside their corresponding connection (frequent/rare). The most notable finding was that the P300 could be explained by an increase

in lateral connection strength (highlighted in bold). Lower panel: The channel positions (left) and ERPs (right) averaged over two subsets of channels (circled on the left). Note the correspondence between the measured ERPs and those generated by the model. See [74] for details of the model and the experiment. In brief, auditory stimuli, 1000 or 2000 Hz tones with 5 ms rise and fall times and 80 ms duration, were presented binaurally. The tones were presented for 15 minutes, every 2 seconds in a pseudo-random sequence with 2000-Hz tones occurring 20% of the time and 1000-Hz tones occurring 80% of the time. The subject was instructed to keep a mental record of the number of 2000-Hz tones (rare target tones). Data were acquired using 128 EEG electrodes with 1000 Hz sample frequency. Before averaging, data were referenced to mean earlobe activity and band-pass filtered between 1 and 30 Hz. Trials showing ocular artefacts and bad channels were removed from further analysis.

UKPMC Funders Group Author Manuscript